



# Systemic infinitesimal over-dispersion on graphical dynamic models

Ning Ning<sup>1</sup> · Edward Ionides<sup>2</sup>

Received: 9 October 2023 / Accepted: 28 May 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

## Abstract

Stochastic models for collections of interacting populations have crucial roles in many scientific fields such as epidemiology, ecology, performance engineering, and queueing theory, to name a few. However, the standard approach to extending an ordinary differential equation model to a Markov chain does not have sufficient flexibility in the mean-variance relationship to match data. To handle that, we develop new approaches using Dirichlet noise to construct collections of independent or dependent noise processes. This permits the modeling of high-frequency variation in transition rates both within and between the populations under study. Our theory is developed in a general framework of time-inhomogeneous Markov processes equipped with a general graphical structure. We demonstrate our approach on a widely analyzed measles dataset, adding Dirichlet noise to a classical Susceptible–Exposed–Infected–Recovered model. Our methodology shows improved statistical fit measured by log-likelihood and provides new insights into the dynamics of this biological system.

**Keywords** Time-inhomogeneous stochastic processes · Directed graph · Extra-demographic stochasticity · Simultaneous jumps · Data science

**Mathematics Subject Classification** 60G20 · 60J20 · 60J22

## 1 Introduction

In this section, we first give the background and motivations in Sect. 1.1 and then state our contributions in Sect. 1.3, followed with the organization of the paper in Sect. 1.4.

### 1.1 Background and motivations

In this paper, we consider a general stochastic graphical dynamic model (GDM). Recalling that a dynamic model is a process whose state varies with time, a GDM is an interacting dynamic model equipped with a graphical structure, where there is a process associated with each vertex whose state varies with time and the states of other vertices. GDMs have wide applications in demography, queueing theory, performance engineering, epidemiology, biology, and other areas, whose examples include stochastic com-

partmental models used in population dynamics. However, mechanistically-inspired models of probabilistic evolution frequently do not contain sufficient variability to adequately match real-world data and further flexibility is still often required (Ramsay and Hooker (2017), page 15). This provides motivation to bestow the desired flexibility on Markov counting processes (MCPs), which are usually used as the building blocks of GDMs. For example, in epidemiology, the conceptual, theoretical, and computational convenience of MCPs has led to their widespread use for modeling disease transmission processes with stochastic compartment models, such as the Susceptible–Exposed–Infected–Recovered (SEIR) model and its generalizations.

When there is at most one event taking place in a sufficiently short period, a MCP is called simple otherwise it is called compound. Bretó and Ionides (2011) showed that infinitesimal dispersion is an equivalent mathematical terminology: a MCP is said to have infinitesimal equi-dispersion (IED) if and only if it is simple and a MCP is said to have infinitesimal over-dispersion (IOD) if and only if it is compound. Using the ratio-formed formula of infinitesimal dispersion, the variance function divided by the mean function of the MCP in a sufficiently short period, IED (resp. IOD,

✉ Ning Ning  
patning@tamu.edu

<sup>1</sup> Department of Statistics, Texas A&M University, College Station, USA

<sup>2</sup> Department of Statistics, University of Michigan, Ann Arbor, USA

infinitesimal under-dispersion (IUD)) holds if the ratio = (resp.  $>$ ,  $<$ ) one. For example, the Poisson process has IED. There are two classes of motivations for modeling IOD. First, the process in question indeed has such occurrences, such as the ruin model in Albrecher et al. (2017) that allows for multiple insurance claims to occur simultaneously (a phenomenon known as *clusters* in actuarial science terminology). Second, in data analysis, we may have multiple event times that are short compared to the scale of primary interest. For example, New York state ([health.data.ny.gov](https://health.data.ny.gov)) has published daily estimates of the number of total COVID-19 tests conducted at (possibly) different time within the day.

By appending gamma noise to constant transition rates, Bretó and Ionides (2011) proposed an IOD generation approach based on simple MCPs. Bretó (2012) provided a multivariate extension for some univariate MCPs considered in Bretó and Ionides (2011) with time-homogeneous transition rate functions (TRFs). Zhang and Li (2016) gave characterizations of discrete compound Poisson distributions and an application in probabilistic number theory. Sendova and Minkova (2018) introduced a compound Poisson counting process with logarithmic compounded distribution. Li and Sendova (2020) proposed a surplus process involving a compound Poisson counting process. The concept of simultaneous co-jumps was proposed in Bretó (2021) with time-homogeneous TRFs. Gao and Sendova (2022) proposed a generalization of the classical compound Poisson model with claim sizes following a compound distribution. There is also similar interest in queueing theory, for instance, the batch Markovian arrival process, which extends the Markovian arrival process by allowing multiple events to occur simultaneously (see, e.g., Maraghi et al. 2009; Jayaraman and Matis 2010).

The assumption of constant or time-homogeneous transition rates is often unrealistic (Krak et al. 2017). Based on the fundamental time-inhomogeneous birth process discussed in Chapter 7 of Klugman et al. (2013), a time-inhomogeneous compound-birth process was recently proposed by Sendova and Minkova (2020). A further generalization from being time-inhomogeneous, is allowing the TRF of one MCP to also depend on the state of others, which is called interacting particle systems in mathematics terminology. In finance, the price of one asset usually depend on time and prices of other assets (e.g. equation (2.7) of Ding and Ning (2021) and equation (2.1) of Ning and Wu (2021)). In epidemiology, the TRF of one compartment in stochastic compartment models usually depend on time and states of other compartments, for example, the rate of new infections in the SEIR-typed Markov chain model (equation (6) on page 332 of Bretó et al. (2009) and equation (1) of this paper). With this kind of general TRFs, Bretó et al. (2009) developed the first over-dispersion methodology for real epidemic data fitting, and this approach has been widely used. Although their TRFs

are quite general, their theoretical foundation is limited to the IOD generation approach on constant rates proposed in Bretó and Ionides (2011).

The long-standing gap between the models used in practice and the theory provided by Bretó et al. (2009) and Bretó and Ionides (2011) is hard to fill. Two natural questions arise: Can an algorithmic approach be developed that performs comparably or better than that of Bretó et al. (2009) without a theory-practice gap? Is this new approach applicable in practice and compatible with modern likelihood-based inference methodologies (e.g. Ionides et al. (2006, 2015); King et al. (2008)) to fully replace that of Bretó et al. (2009)? Graphs, as a kind of data structure that models a set of objects and their relationships, can be used as a denotation of a large number of systems across various areas. Because of their great expressive power, researches on analyzing GDMs systemically have been receiving more and more attentions in many areas, Chen et al. (2017) on network reconstruction from high-dimensional ordinary differential equations, Katzfuss et al. (2020) on ensemble Kalman methods for high-dimensional hierarchical dynamic space-time models, Ionides et al. (2021) on bagged filters for partially observed spatiotemporal systems, Ning and Ionides (2023) on high-dimensional spatiotemporal online learning on large graphs, to name a few. Then two more fundamental questions arise directly: Can a systemic theory be defined properly and established rigorously on a general graph instead of merely on edges? What sort of mathematical tools are needed to build a systemic theory to be exploited algorithmically? In this paper, we aim to address the above four questions.

## 1.2 Motivational application

We use measles modeling as a motivational example here, whose real data analysis is provided in Sect. 4. Worldwide, measles remains a leading cause of vaccine-preventable death and disability, however global eradication of this highly infectious disease by intensive vaccination would be difficult. A fundamental class of models for measles transmission is the SEIR model, where (S) represents susceptible individuals who have not been infected yet but may experience infection later, (E) represents individuals exposed and carrying a latent infection, (I) represents infectious individuals that have been infected and are infectious to others, and (R) represents recovered individuals that are no longer infectious and are immune. Two other compartments/vertices (B) and (D) representing the birth and death of individuals respectively, are added in SEIR-type Markov chain models which have been commonly used for measles data analysis. The directed graph in Fig. 1 gives a diagrammatic representation, where arrows are used to indicate the possibility of transitions between vertices with labels parameterizing the transition rates.

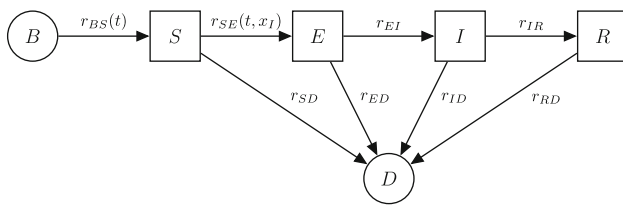


Fig. 1 Directed graph for SEIR-type Markov chain models

The state of the system at time  $t$  is given by the number of individuals in each vertex and is denoted as

$$\mathbf{X}(t) = \mathbf{x} = \{x_v\}_{v \in V} = \{x_B, x_S, x_E, x_I, x_R, x_D\},$$

where  $V = \{B, S, E, I, R, D\}$ . According to Bretó et al. (2009) and He et al. (2010), the TRF of infection is specified as

$$r_{SE}(t, x_I) = \beta(t)(x_I + \iota)^\alpha / P(t), \tag{1}$$

where  $\beta(t)$  is the transmission rate defined in (20),  $\iota$  describes imported infectives,  $\alpha$  is a mixing parameter with  $\alpha = 1$  corresponding to homogeneous mixing, and  $P(t)$  is a known population size obtained via interpolation from census data. We can see that the transition rate over arrow  $(S, E)$  depends on time  $t$  which makes the dynamic time-inhomogeneous, and depends on the value  $x_I$  of vertex  $I$  which necessitates systemic graphical analysis.

### 1.3 Our contributions

In this paper, we define systemic IOD (SIOD) for general GDMs, provide corresponding methodologies for general dynamics, generate associated general algorithms, and demonstrate the algorithmic performance on a benchmark epidemiological modeling challenge. In sum, the contributions of this paper are four-fold:

- (1) *General GDM and systemic definitions.* The GDM under consideration is general in terms of a general graph structure and general dynamics over it. We consider TRFs as general positive functions of time and the state of the whole graph, while all the preceding IOD theoretical literature considered either constants or functions of time only. We focus on dynamics over a general directed graph, while all the preceding IOD theoretical literature worked on dynamics over a single arrow of the graph. We hence give appropriate definitions of systemic infinitesimal dispersion (SID), which allow users to flexibly add IOD to dynamics over some subgraphs while keeping dynamics over the rest subgraphs having IED.
- (2) *Innovative methodologies and algorithms.* On one hand, under boundedness constraints, in Sect. 3.1 we gener-

ate IOD using multinomial distributions, over outgoing arrows with the same tail. An algorithmic Euler realization of the resulting Theorem 3.1 is provided, which is a general algorithm (Algorithm 1) for generating dynamics having IOD over connected outgoing arrows. Its application to a well-known case study in epidemiology is provided in Algorithm 2. On the other hand, without boundedness constraints, in Sect. 3.2 we propose a methodology for generating IOD using negative multinomial distributions, over incoming arrows with the same head and a corresponding general algorithm can be developed analogously.

- (3) *Wide applicability.* Our theoretical framework and methodologies are sufficiently general to cover many situations. First, only a weak assumption is required (existence of the second moment of a single dynamic), which is usually satisfied in practice; second, software implementation using Dirichlet random variables is routine and computationally convenient; third, with our definition of SIOD, users can flexibly choose those subgraphs that are appropriate to apply our algorithm for overdispersion; fourth, the convenience of simulation from the proposed algorithms enables likelihood-based data fitting using simulation-based algorithms, among which we demonstrated using iterated filtering (Ionides et al. 2011, 2006, 2015; King et al. 2008); fifth, just one additional parameter is needed to describe overdispersion and this parameter can be inferred using these aforementioned simulation-based algorithms (Sect. 4.2).
- (4) *Improved data fitting with new insights.* Besides the long-standing theory-practice gap, there are long-existing concerns about interpreting results. He et al. (2010) applied the algorithm proposed in Bretó et al. (2009) on a benchmark epidemiological modeling challenge. They obtained a surprisingly large  $R_0$  value, which is the basic reproduction number that is central in epidemiological theory. In Sect. 4.2, we conduct fair comparisons by applying our algorithm with the same data, same model setting, and same inference algorithm. We achieve better data fitting and provide a resolution of a previous discrepancy on maximum likelihood estimation (MLE).

### 1.4 Organization of the paper

The rest of the paper proceeds as follows. In Sect. 2, we give the graph structure, configurations on the graph, dynamics over the graph, and definitions of infinitesimal dispersion. Our methodology for generating SIOD is provided in Sect. 3, where Sects. 3.1 and 3.2 cover dynamics with and without boundedness constraints, respectively. In Sect. 4, we first describe a SEIR-type Markov chain model which has been commonly used for measles data analysis in Sect. 4.1, and then conduct measles real data analysis on a well-tested and pub-

licly accessible dataset using our proposed IOD methodology and algorithm in Sect. 4.2. We conclude with discussion in Sect. 5. Proofs of the paper are provided in Section A1 of the Appendix. Code and data reproducing our results are available online at <https://github.com/patning/Over-dispersion>. The notations used throughout this paper are listed in Table 3.

## 2 Time-inhomogeneous GDMs

In this section, we first give the model description in Sect. 2.1, and then definitions of infinitesimal dispersion in Sect. 2.2.

### 2.1 Model setup

In this paper, a general GDM is formally defined by two components: the first is a general directed graph; the second is stochastic processes having general TRFs associated with each vertex of the graph.

A directed graph is a set of vertices connected by edges, where each edge has a direction associated with it. In this paper, we consider a finite directed graph as  $G = (V, A)$ , where  $V$  is a set of vertices and  $A$  is a set of arrows. For the directed graph in Fig. 1,  $V = \{B, S, E, I, R, D\}$  and

$$A = \{(B, S), (S, E), (E, I), (I, R), (S, D), (E, D), (I, D), (R, D)\}.$$

For two vertices  $v, v' \in V$ , an arrow  $(v, v')$  is considered to be directed from  $v$  to  $v'$ ;  $v'$  is called the head and  $v$  is called the tail of the arrow. In this paper, we allow the directed graph to have loops, i.e., arrows that directly connect vertices with themselves. For a vertex  $v \in V$ , the number of head ends adjacent to  $v$  is called the indegree of  $v$  and is denoted as  $\text{deg}^-(v)$ ; the number of tail ends adjacent to  $v$  is called the outdegree of  $v$  and is denoted as  $\text{deg}^+(v)$ . A vertex with zero indegree is called a source and the set of all source vertices is denoted by  $S_o$ . A vertex with zero outdegree is called a sink and the set of all sink vertices is denoted by  $S_i$ . Thus,

$$S_o := \{v \in V; \text{deg}^-(v) = 0\}$$

and  $S_i := \{v \in V; \text{deg}^+(v) = 0\}.$

For the directed graph in Fig. 1,  $S_o = \{B\}$  and  $S_i = \{D\}$ . Denote the set of incoming neighbors of  $v$  as  $N_G^-(v)$ , which is the set of vertices  $\bar{v} \in A$  such that  $(\bar{v}, v) \in A$ . Denote the set of outgoing neighbors of  $v$  as  $N_G^+(v)$ , which is the set of vertices  $v' \in A$  such that  $(v, v') \in A$ . Directed graphs distinguish between  $N_G^-(v)$  and  $N_G^+(v)$ .

Given a Polish space  $\mathcal{H}$ , we let  $\mathcal{D}_{\mathcal{H}}[0, \infty)$  denote the space of  $\mathcal{H}$ -valued càdlàg functions on  $[0, \infty)$ , endowed with the Skorokhod  $J_1$  topology, such that  $\mathcal{D}_{\mathcal{H}}[0, \infty)$  is a Polish

space; see Parthasarathy (2005) for further theoretical details. Denote the random variable (also known as spin in graph theory terminology) on any vertex  $v \in V$  at any time  $t \in [0, \infty)$  as  $X_v(t)$ . We consider that  $X_v(t)$  is defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and takes values in  $\mathcal{D}_{\mathcal{H}}[0, \infty)$ , equipped with the Borel  $\sigma$ -algebra generated by open sets under the Skorokhod  $J_1$ -topology. The number of transitions from vertex  $v$  to vertex  $v'$  through arrow  $(v, v') \in A$  is modeled by a nondecreasing integer-valued jump process  $N_{vv'}^{\mathbf{X}}(t)$  for  $t \in [0, \infty)$  defined on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , where we use the customary initialization  $N_{vv'}^{\mathbf{X}}(0) = 0$ .

Suppose that the dynamics of  $\mathbf{X}(t) := \{X_v(t)\}_{v \in V}$  are driven by  $\mathbf{N}^{\mathbf{X}}(t) := \{N_{vv'}^{\mathbf{X}}(t)\}_{(v,v') \in A}$  as follows: For any  $v \in V$  and  $t \in [0, \infty)$

$$X_v(t) = X_v(0) + \sum_{\bar{v} \in N_G^-(v)} N_{\bar{v}v}^{\mathbf{X}}(t) - \sum_{v' \in N_G^+(v)} N_{vv'}^{\mathbf{X}}(t).$$

That is, the random variable of vertex  $v$  at time  $t$  is given by its initial value at time 0, plus the increments from all its incoming neighbors, and then minus the decrements to its outgoing neighbors. Denote the increment in the time interval  $[t, t + h]$  over arrow  $(v, v') \in A$  as

$$\Delta_{vv'}^{\mathbf{X}}(t, h) := N_{vv'}^{\mathbf{X}}(t + h) - N_{vv'}^{\mathbf{X}}(t),$$

and then

$$X_v(t + h) - X_v(t) = \sum_{\bar{v} \in N_G^-(v)} \Delta_{\bar{v}v}^{\mathbf{X}}(t, h) - \sum_{v' \in N_G^+(v)} \Delta_{vv'}^{\mathbf{X}}(t, h).$$

### 2.2 Measures of dispersion

Measures of dispersion were defined previously in the variance to mean ratio form (e.g. Gillespie 1984) and the variance and mean difference form (e.g. Brown et al. 1998). For theoretical analysis of dispersion, these two kinds of definitions are mainly equivalent while the difference-formed definition avoids the “0/0” situation. However, the ratio-formed definition is widely used, partially due to the fact that it facilitates the dispersion comparison among different metrics and/or units. When it comes to data analysis, the over-dispersion parameter in Poisson regression (see, e.g., Berk and MacDonald 2008) uses the ratio-formed definition. In this paper, we properly define the SID with respect to the whole graph in Definition 2.2, which is the first time the measure of dispersion is defined on a graph to our best knowledge. Definition 2.2 is formulated in terms of the measure of dispersion with respect to each arrow of the graph, whose definition is given below and is consistent with that in Bretó and Ionides (2011) (equation (3) on page 2574):

**Definition 2.1** For arrow  $(v, v') \in A$ , define the infinitesimal variance

$$[\sigma_{vv'}^{d\mathbf{X}}(t, \mathbf{x})]^2 := \lim_{h \downarrow 0} h^{-1} \text{Var}[\Delta_{vv'}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}],$$

and the infinitesimal mean

$$\mu_{vv'}^{d\mathbf{X}}(t, \mathbf{x}) := \lim_{h \downarrow 0} h^{-1} \mathbb{E}[\Delta_{vv'}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}].$$

Define the infinitesimal dispersion index as the following ratio if it exists:

$$D_{vv'}^{d\mathbf{X}}(t, \mathbf{x}) := [\sigma_{vv'}^{d\mathbf{X}}(t, \mathbf{x})]^2 / \mu_{vv'}^{d\mathbf{X}}(t, \mathbf{x}).$$

We say that with respect to arrow  $(v, v')$ ,  $\mathbf{X}(t)$  has IED at  $\mathbf{X}(t) = \mathbf{x}$  if  $D_{vv'}^{d\mathbf{X}}(t, \mathbf{x}) = 1$ , has IOD at  $\mathbf{X}(t) = \mathbf{x}$  if  $D_{vv'}^{d\mathbf{X}}(t, \mathbf{x}) > 1$ , and has IUD at  $\mathbf{X}(t) = \mathbf{x}$  if  $D_{vv'}^{d\mathbf{X}}(t, \mathbf{x}) < 1$ .

Noting that Definition 2.1 is with respect to a specific arrow, now we give definitions with respect to the whole graph. A GDM having SIOD is provided in Sect. 4.

**Definition 2.2** We say that

- $\mathbf{X}(t)$  has SIED at  $\mathbf{X}(t) = \mathbf{x}$ , if  $D_{vv'}^{d\mathbf{X}}(t, \mathbf{x}) = 1$  for all  $(v, v') \in A$ ;
- $\mathbf{X}(t)$  has SIOD at  $\mathbf{X}(t) = \mathbf{x}$ , if  $D_{vv'}^{d\mathbf{X}}(t, \mathbf{x}) \geq 1$  for all  $(v, v') \in A$  and there exists  $(v_0, v'_0) \in A$  such that  $D_{v_0 v'_0}^{d\mathbf{X}}(t, \mathbf{x}) > 1$ ;
- $\mathbf{X}(t)$  has SIUD at  $\mathbf{X}(t) = \mathbf{x}$ , if  $D_{vv'}^{d\mathbf{X}}(t, \mathbf{x}) \leq 1$  for all  $(v, v') \in A$  and there exists  $(v_0, v'_0) \in A$  such that  $D_{v_0 v'_0}^{d\mathbf{X}}(t, \mathbf{x}) < 1$ .

Note that the above definitions depend on arrow-wise variances. To explore the infinitesimal correlations between two arrows' dynamics, in the following we give the pairwise definition of infinitesimal covariance consistently with the arrow-wise definition of infinitesimal variance in Definition 2.1.

**Definition 2.3** For arrows  $(u, u') \in A$  and  $(v, v') \in A$ , define the infinitesimal covariance

$$\begin{aligned} \sigma_{uu', vv'}^{d\mathbf{X}}(t, \mathbf{x}) \\ := \lim_{h \downarrow 0} h^{-1} \text{Cov}[\Delta_{uu'}^{\mathbf{X}}(t, h), \Delta_{vv'}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}]. \end{aligned}$$

### 3 Probabilistic construction of IOD

In this section, we aim to generate a new model  $\mathbf{X}$  having SIOD based on a GDM  $\mathbf{Z}$  having SIED. We consider  $\mathbf{Z}$  in a general form in the way that conditional on  $\mathbf{Z}(t) = \mathbf{z}$ , each

flow  $N_{vv'}^{\mathbf{Z}}$  over arrow  $(v, v') \in A$  is associated with a general TRF  $\Upsilon_{vv'}$  which depends on time  $(t)$  and state of the graph  $(\mathbf{z})$ , such that

$$\begin{aligned} Q(t, (z_v, z_{v'}, \mathbf{z}_{V \setminus \{v, v'\}}), (z_v - 1, z_{v'} + 1, \mathbf{z}_{V \setminus \{v, v'\}})) \\ = \Upsilon_{vv'}(t, \mathbf{z}), \end{aligned} \tag{2}$$

where we used the standard definition of the transition rate

$$Q(t, \mathbf{x}, \mathbf{x}') := \lim_{h \downarrow 0} h^{-1} \mathbb{P}(\mathbf{X}(t+h) = \mathbf{x}' \mid \mathbf{X}(t) = \mathbf{x}). \tag{3}$$

The Markov chain interpretation of  $\mathbf{Z}$  can be specified by the infinitesimal transition probabilities:

$$\begin{aligned} \mathbb{P}(\Delta_{vv'}^{\mathbf{Z}}(t, h) = 0 \mid \mathbf{Z}(t) = \mathbf{z}) &= 1 - \Upsilon_{vv'}(t, \mathbf{z})h + o(h), \\ \mathbb{P}(\Delta_{vv'}^{\mathbf{Z}}(t, h) = 1 \mid \mathbf{Z}(t) = \mathbf{z}) &= \Upsilon_{vv'}(t, \mathbf{z})h + o(h), \\ \mathbb{P}(\Delta_{vv'}^{\mathbf{Z}}(t, h) > 1 \mid \mathbf{Z}(t) = \mathbf{z}) &= o(h), \\ \mathbb{P}(\Delta_{vv'}^{\mathbf{Z}}(t, h) < 0 \mid \mathbf{Z}(t) = \mathbf{z}) &= 0. \end{aligned} \tag{4}$$

Without loss of generality, we suppose the initial values of the dynamics over the graph,  $\mathbf{Z}(0)$ , are integers for notational simplicity. In Sects. 3.1 and 3.2, we consider the transition rate (defined in (2)) in the forms given in (5) and (14), respectively.

#### 3.1 IOD construction with boundedness constraints

In this subsection, we focus on generating GDMs having SIOD over outgoing arrows with the same tail. We consider the case that there are multiple connected outgoing arrows of vertex  $v$  such that  $|N_G^+(v)| \geq 1$ , where  $N_G^+(v)$  is the set of vertices  $v' \in V$  such that  $(v, v') \in A$  and  $|N_G^+(v)|$  is its cardinality. Suppose  $N_G^+(v) = \{v'_1, \dots, v'_m\}$  and  $|N_G^+(v)| = m$ . For example in Fig. 1, the connected outgoing arrows of vertex  $S$  are  $(S, E)$  and  $(S, D)$ , which gives  $N_G^+(S) = \{E, D\}$  and  $m = 2$ .

The transition rate of  $N_{vv'_i}^{\mathbf{Z}}(t)$  for  $i \in \{1, \dots, m\}$  is given by

$$\begin{aligned} Q(t, (z_v, z_{v'_i}, \mathbf{z}_{V \setminus \{v, v'_i\}}), (z_v - 1, z_{v'_i} + 1, \mathbf{z}_{V \setminus \{v, v'_i\}})) \\ = r_{vv'_i}(t, \mathbf{z}) z_v \mathbb{1}_{\{z_v \geq 1\}}, \end{aligned} \tag{5}$$

where  $\mathbf{z} = (z_v, z_{v'_i}, \mathbf{z}_{V \setminus \{v, v'_i\}})$ . Considering the time period  $[t, t+h]$ , by (4) the probability that one transition from vertex  $v$  to vertex  $v'_i$  for  $i \in \{1, \dots, m\}$  is given by

$$\mathbb{P}(\Delta_{vv'_i}^{\mathbf{Z}}(t, h) = 1 \mid \mathbf{Z}(t) = \mathbf{z}) = r_{vv'_i}(t, \mathbf{z}) z_v h + o(h).$$

For notational convenience, denote  $\Delta_{vv'_0}^{\mathbf{Z}}(t, h)$  as the remaining individuals at vertex  $v$ . Then the joint distribution of

$\{\Delta_{vv'_i}^{\mathbf{Z}}(t, h) = k_i\}_{i \in \{0, \dots, m\}}$  is given by

$$\mathbb{P}\left(\{\Delta_{vv'_i}^{\mathbf{Z}}(t, h) = k_i\}_{i \in \{0, \dots, m\}} \mid \mathbf{Z}(t) = \mathbf{z}\right) = \frac{\Gamma(z_v + 1)}{\prod_{i=0}^m \Gamma(k_i + 1)} \prod_{i=0}^m \left[\tilde{\pi}_{vv'_i}(t, h, \mathbf{z})\right]^{k_i} + o(h), \tag{6}$$

where  $\Gamma(\cdot)$  is the gamma function,  $z_v \geq 1$  and  $k_i \in \{0, 1, \dots, z_v\}$  for  $i \in \{0, \dots, m\}$  such that  $\sum_{i=0}^m k_i = z_v$ . Here, for  $i \in \{1, \dots, m\}$

$$\begin{aligned} \tilde{\pi}_{vv'_i}(t, h, \mathbf{z}) &= \left(1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{z}) ds}\right) \\ &\times \frac{r_{vv'_i}(t, \mathbf{z})z_v h + o(h)}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{z})z_v h + o(h)} + o(h) \\ &= \left(1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{z}) ds}\right) \\ &\times \frac{r_{vv'_i}(t, \mathbf{z})z_v h}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{z})z_v h + o(h)} + o(h) \\ &= \left(1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{z}) ds}\right) \\ &\times \frac{r_{vv'_i}(t, \mathbf{z})z_v h}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{z})z_v h} \left(\frac{1}{1 + o(h)}\right) + o(h) \\ &= \left(1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{z}) ds}\right) \\ &\times \frac{r_{vv'_i}(t, \mathbf{z})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{z})} (1 + o(h)) + o(h) \\ &= \left(1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{z}) ds}\right) \\ &\times \frac{r_{vv'_i}(t, \mathbf{z})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{z})} + o(h), \tag{7} \end{aligned}$$

where we used Taylor series in the fourth equality, and

$$\tilde{\pi}_{vv'_0}(t, h, \mathbf{z}) = 1 - \sum_{i=1}^m \tilde{\pi}_{vv'_i}(t, h, \mathbf{z}).$$

Plugging (7) into (6), we can rewrite (6) as

$$\mathbb{P}\left(\{\Delta_{vv'_i}^{\mathbf{Z}}(t, h) = k_i\}_{i \in \{0, \dots, m\}} \mid \mathbf{Z}(t) = \mathbf{z}\right) = \frac{\Gamma(z_v + 1)}{\prod_{i=0}^m \Gamma(k_i + 1)} \prod_{i=0}^m \left[\pi_{vv'_i}(t, h, \mathbf{z})\right]^{k_i} + o(h), \tag{8}$$

where

$$\pi_{vv'_i}(t, h, \mathbf{z})$$

$$= \begin{cases} \left(1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{z}) ds}\right) \frac{r_{vv'_i}(t, \mathbf{z})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{z})} & i \in \{1, \dots, m\}, \\ 1 - \sum_{j=1}^m \pi_{vv'_j}(t, h, \mathbf{z}) & i = 0. \end{cases} \tag{A.1}$$

Denote the transition rate  $q_{vv'}$  at time  $t$  over arrow  $(v, v') \in A$  as

$$q_{vv'}(t, \mathbf{x}, k) := \lim_{h \downarrow 0} h^{-1} \mathbb{P}(\Delta_{vv'}^{\mathbf{X}}(t, h) = k \mid \mathbf{X}(t) = \mathbf{x}). \tag{9}$$

The following theorem shows that a GDM  $\mathbf{X}$  having SIOD can be generated over connected outgoing arrows  $\{(v, v'_i)\}_{i \in \{1, \dots, m\}}$ .

**Theorem 3.1** *Suppose that  $r_{vv'_i}(t, \mathbf{x})$ , for each  $i \in \{1, \dots, m\}$ , is a positive function that is uniformly continuous in  $t$ . Further suppose that  $\{\Delta_{vv'_i}^{\mathbf{X}}(t, h) = k_i\}_{i \in \{0, \dots, m\}}$  are jointly distributed in the time period  $[t, t + h]$  as follows:*

$$\begin{aligned} \mathbb{P}\left(\{\Delta_{vv'_i}^{\mathbf{X}}(t, h) = k_i\}_{i \in \{0, \dots, m\}} \mid \mathbf{X}(t) = \mathbf{x}, \{\Pi_{vv'_i}(t, h, \mathbf{x})\}_{i \in \{0, \dots, m\}}\right) \\ = \frac{\Gamma(x_v + 1)}{\prod_{i=0}^m \Gamma(k_i + 1)} \prod_{i=0}^m \left(\Pi_{vv'_i}(t, h, \mathbf{x})\right)^{k_i} + o(h), \tag{10} \end{aligned}$$

where  $x_v \geq 1$  and  $k_i \in \{0, 1, \dots, x_v\}$  for  $i \in \{0, \dots, m\}$  such that  $\sum_{i=0}^m k_i = x_v$ . Further suppose that the family  $\{\Pi_{vv'_i}(t, h, \mathbf{x})\}_{i \in \{0, 1, \dots, m\}}$  is distributed according to the Dirichlet distribution  $\text{Dir}(\{\alpha_{vv'_i}(t, h, \mathbf{x})\}_{i \in \{0, 1, \dots, m\}})$  having

$$\alpha_{vv'_i}(t, h, \mathbf{x}) = c \pi_{vv'_i}(t, h, \mathbf{x}) \quad \text{for } i \in \{0, \dots, m\},$$

where  $c$  is a finite positive parameter and

$$\pi_{vv'_i}(t, h, \mathbf{x}) = \begin{cases} \left(1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{x}) ds}\right) \frac{r_{vv'_i}(t, \mathbf{x})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{x})} & i \in \{1, \dots, m\}, \\ 1 - \sum_{j=1}^m \pi_{vv'_j}(t, h, \mathbf{x}) & i = 0. \end{cases}$$

The following results hold:

(1) For each  $i \in \{1, \dots, m\}$ , the infinitesimal mean  $\mu_{vv'_i}^{d\mathbf{X}}(t, \mathbf{x})$  is given by

$$\mu_{vv'_i}^{d\mathbf{X}}(t, \mathbf{x}) = x_v r_{vv'_i}(t, \mathbf{x})$$

and the infinitesimal variance  $[\sigma_{vv'_i}^{d\mathbf{X}}(t, \mathbf{x})]^2$  is given by

$$[\sigma_{vv'_i}^{d\mathbf{X}}(t, \mathbf{x})]^2 = (1 + (x_v - 1)(c + 1)^{-1})x_v r_{vv'_i}(t, \mathbf{x}).$$

When  $x_v > 1$ ,  $\mathbf{X}(t)$  has IOD at  $\mathbf{X}(t) = \mathbf{x}$  with respect to each arrow of  $\{(v, v'_i)\}_{i \in \{1, \dots, m\}}$  and  $\mathbf{X}(t)$  has SIOD at  $\mathbf{X}(t) = \mathbf{x}$  for connected outgoing arrows  $\{(v, v'_i)\}_{i \in \{1, \dots, m\}}$ ; when  $x_v = 1$ ,  $\mathbf{X}(t)$  has SIED at  $\mathbf{X}(t) = \mathbf{x}$  for connected outgoing arrows  $\{(v, v'_i)\}_{i \in \{1, \dots, m\}}$ . Furthermore, for  $i, j \in \{1, \dots, m\}$  and  $i \neq j$ , the infinitesimal covariance  $\sigma_{vv'_i, vv'_j}^{d\mathbf{X}}(t, \mathbf{x}) = 0$ .

(2) Denote  $\mathcal{S}$  as the set of transitions over arrows  $\{(v, v_i)\}_{i \in \{1, \dots, m\}}$ , i.e.,

$$\mathcal{S} := \left\{ k_i : k_i \geq 1 \text{ for } i \in \{1, \dots, m\} \text{ and } \sum_{i=0}^m k_i = x_v \right\}. \tag{11}$$

Then the conditional probability that transitions happen over two or more arrows

$$\mathbb{P}\left(\{\Delta_{vv'_i}^{\mathbf{X}}(t, h) = k_i\}_{i \in \{0, \dots, m\}}, |\mathcal{S}| \geq 2 \mid \mathbf{X}(t) = \mathbf{x}\right) = o(h)$$

and the conditional probability that only one transition happens over a single arrow

$$\mathbb{P}\left(\{\Delta_{vv'_i}^{\mathbf{X}}(t, h) = k_i\}_{i \in \{0, \dots, m\}}, |\mathcal{S}| = 1 \mid \mathbf{X}(t) = \mathbf{x}\right) = \sum_{i=1}^m q_{vv'_i}(t, \mathbf{x}, k_i)h + o(h), \tag{12}$$

where  $|\mathcal{S}|$  is the cardinality of  $\mathcal{S}$  and for  $i \in \{1, \dots, m\}$

$$q_{vv'_i}(t, \mathbf{x}, k_i) = c \binom{x_v}{k_i} \frac{\Gamma(k_i)\Gamma(x_v - k_i + c)}{\Gamma(x_v + c)} r_{vv'_i}(t, \mathbf{x}). \tag{13}$$

The proof of Theorem 3.1 is postponed to Section A1.1 in the Appendix. We note that a crucial difference between equations (4) and (12) is that, one transition over any single arrow has up to  $m (\geq 1)$  units in (12) while one transition over a specific arrow has exactly one unit in (4). From Theorem 3.1, we can see that the methodology proposed in Bretó and Ionides (2011) is even a special case of our  $m = 1$  case. Now, we realize the methodology proposed in Theorem 3.1 in the Algorithm 1.

**Algorithm 1** Euler scheme on generating dynamics having IOD over connected outgoing arrows  $\{(v, v'_i)\}_{i \in \{1, \dots, m\}}$ .

```

Set the initial value  $\mathbf{X}(0)$  and time interval  $[0, T]$ .
Set fixed time increment  $\delta = T/N$  for integer  $N$ ; define  $t_n = n\delta$ .
FOR  $n = 0$  to  $N - 1$ 
  FOR each  $v \in V$  with  $N_G^+(v) = \{v'_1, \dots, v'_m\}$ 
    Generate  $\{\Pi_{vv'_i}\}_{i \in \{0, 1, \dots, m\}}$  according to the Dirichlet distribution
    Dir( $\{\alpha_{vv'_i}\}_{i \in \{0, 1, \dots, m\}}$ ) having
     $\alpha_{vv'_i} = c\pi_{vv'_i}$  for  $i \in \{1, \dots, m\}$  and  $\alpha_{vv'_0} = c - \sum_{i=1}^m \alpha_{vv'_i}$ ,
    where for  $i \in \{1, \dots, m\}$ 
     $\pi_{vv'_i} = \left(1 - e^{-\sum_{j=1}^m \int_{t_n}^{t_{n+1}} r_{vv'_j}(s, \mathbf{x}) ds}\right) \frac{r_{vv'_i}(t_n, \mathbf{x})}{\sum_{j=1}^m r_{vv'_j}(t_n, \mathbf{x})}$ .
    Generate process increments
     $\{\Delta_{vv'_i}^{\mathbf{X}}\}_{i \in \{0, 1, \dots, m\}} \sim \text{Multinomial}(X_v(t_n), \{\Pi_{vv'_i}\}_{i \in \{0, 1, \dots, m\}})$ 
    where  $\Delta_{vv'_0}^{\mathbf{X}}$  stands for retain individuals.
    Set  $X_v(t_{n+1}) = \Delta_{vv'_0}^{\mathbf{X}} + \sum_{\bar{v} \in N_G^-(v)} \Delta_{\bar{v}v}^{\mathbf{X}}$ 
  END FOR
END FOR
    
```

**3.2 IOD construction without boundedness constraints**

Unbounded processes, such as the pure birth process, have wide applications. In this subsection, we focus on generating GDMs having IOD over incoming arrows with the same head without boundedness constraints.

We consider the case that there are multiple connected incoming arrows of vertex  $u'$  such that  $|N_G^-(u')| \geq 1$ , where  $N_G^-(u')$  is the set of vertices  $u \in V$  such that  $(u, u') \in A$  and  $|N_G^-(u')|$  is its cardinality. Suppose  $N_G^-(u') := \{u_1, \dots, u_{\bar{m}}\}$  and  $\bar{m} := |N_G^-(u')|$ . For example in Fig. 1, the connected incoming arrows of vertex  $D$  are  $(S, D)$ ,  $(E, D)$ ,  $(I, D)$ , and  $(R, D)$ , which gives  $N_G^-(D) = \{S, E, I, R\}$  and  $\bar{m} = 4$ .

The transition probability of  $N_{u_i u'}^{\mathbf{Z}}(t)$  for  $i \in \{1, \dots, \bar{m}\}$  is given by

$$Q(t, (z_{u_i}, z_{u'}, \mathbf{z}_{V \setminus \{u_i, u'\}}), (z_{u_i} - 1, z_{u'} + 1, \mathbf{z}_{V \setminus \{u_i, u'\}})) = r_{u_i u'}(t, \mathbf{z}) z_{u'} \mathbb{1}_{\{z_{u'} > 0\}}, \tag{14}$$

where  $\mathbf{z} = (z_{u_i}, z_{u'}, \mathbf{z}_{V \setminus \{u_i, u'\}})$ . In the time period  $[t, t + h]$ , by (4) the probability that one transition from vertex  $u_i$  to vertex  $u'$

$$\mathbb{P}(\Delta_{u_i u'}^{\mathbf{Z}}(t, h) = 1 \mid \mathbf{Z}(t) = \mathbf{z}) = r_{u_i u'}(t, \mathbf{z}) z_{u'} h + o(h).$$

The joint distribution of increments of  $\{N_{u_i u'}^{\mathbf{Z}}(t)\}_{i \in \{1, \dots, \bar{m}\}}$  is given by

$$\mathbb{P}\left(\{\Delta_{u_i u'}^{\mathbf{Z}}(t, h) = k_i\}_{i \in \{1, \dots, \bar{m}\}} \mid \mathbf{Z}(t) = \mathbf{z}\right)$$

$$\begin{aligned}
 &= \frac{\Gamma(z_{u'} + \sum_{i=1}^{\bar{m}} k_i)}{\Gamma(z_{u'}) \prod_{i=1}^{\bar{m}} \Gamma(k_i + 1)} \left[ 1 - \sum_{i=1}^{\bar{m}} \tilde{\pi}_{u_i u'}(t, h, \mathbf{z}) \right]^{z_{u'}} \\
 &\times \prod_{i=0}^{\bar{m}} [\tilde{\pi}_{u_i u'}(t, h, \mathbf{z})]^{k_i} + o(h), \tag{15}
 \end{aligned}
 = \begin{cases} \left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{z}) ds} \right) \\ \frac{r_{u_i u'}(t, \mathbf{z})}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{z})} & i \in \{1, \dots, \bar{m}\}, \\ 1 - \sum_{j=1}^{\bar{m}} \pi_{u_j u'}(t, h, \mathbf{z}) & i = 0. \end{cases}$$

where  $z_{u'} > 0$  and  $k_i \in \{0, 1, 2, \dots\}$  for  $i \in \{1, \dots, \bar{m}\}$ . Here, for  $i \in \{1, \dots, \bar{m}\}$

$$\begin{aligned}
 &\tilde{\pi}_{u_i u'}(t, h, \mathbf{z}) \\
 &= \left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{z}) ds} \right) \\
 &\times \frac{r_{v_i v'}(t, \mathbf{z}) z_{u'} h + o(h)}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{z}) z_{u'} h + o(h)} + o(h) \\
 &= \left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{z}) ds} \right) \\
 &\times \frac{r_{v_i v'}(t, \mathbf{z}) z_{u'} h}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{z}) z_{u'} h + o(h)} + o(h) \\
 &= \left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{z}) ds} \right) \\
 &\times \frac{r_{v_i v'}(t, \mathbf{z}) z_{u'} h}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{z}) z_{u'} h} \left( \frac{1}{1 + o(h)} \right) + o(h) \\
 &= \left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{z}) ds} \right) \\
 &\times \frac{r_{v_i v'}(t, \mathbf{z})}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{z})} (1 + o(h)) + o(h) \\
 &= \left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{z}) ds} \right) \\
 &\times \frac{r_{u_i u'}(t, \mathbf{z})}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{z})} + o(h), \tag{16}
 \end{aligned}$$

where we used Taylor series in the fourth equality. Plugging (16) into (15), we can rewrite (15) as

$$\begin{aligned}
 &\mathbb{P}\left(\{\Delta_{u_i u'}^{\mathbf{Z}}(t, h) = k_i\}_{i \in \{1, \dots, \bar{m}\}} \mid \mathbf{Z}(t) = \mathbf{z}\right) \\
 &= \frac{\Gamma(z_{u'} + \sum_{i=1}^{\bar{m}} k_i)}{\Gamma(z_{u'}) \prod_{i=1}^{\bar{m}} \Gamma(k_i + 1)} [\pi_{u_0 u'}(t, h, \mathbf{z})]^{z_{u'}} \\
 &\times \prod_{i=0}^{\bar{m}} [\pi_{u_i u'}(t, h, \mathbf{z})]^{k_i} + o(h),
 \end{aligned}$$

where

$$\pi_{u_i u'}(t, h, \mathbf{z})$$

The following theorem shows that a GDM  $\mathbf{X}$  having SIOD can be generated over connected incoming arrows  $\{(u_i, u')\}_{i \in \{1, \dots, \bar{m}\}}$ .

**Theorem 3.2** Suppose that  $r_{u_i u'}(t, \mathbf{x})$ , for each  $i \in \{1, \dots, \bar{m}\}$ , is a positive function that is uniformly continuous in  $t$ . Further suppose that the increments of  $\{N_{u_i u'}^{\mathbf{X}}(t)\}_{i \in \{1, \dots, \bar{m}\}}$  are jointly distributed in the time period  $[t, t + h]$  as:

$$\begin{aligned}
 &\mathbb{P}\left(\{\Delta_{u_i u'}^{\mathbf{X}}(t, h) = k_i\}_{i \in \{1, \dots, \bar{m}\}} \mid \mathbf{X}(t) \right. \\
 &= \mathbf{x}, \{\Pi_{u_i u'}(t, h, \mathbf{x})\}_{i \in \{0, 1, \dots, \bar{m}\}} \Big) \\
 &= \frac{\Gamma(x_{u'} + \sum_{i=1}^{\bar{m}} k_i)}{\Gamma(x_{u'}) \prod_{i=1}^{\bar{m}} \Gamma(k_i + 1)} \\
 &\times [\Pi_{u_0 u'}(t, h, \mathbf{x})]^{x_{u'}} \prod_{i=1}^{\bar{m}} [\Pi_{u_i u'}(t, h, \mathbf{x})]^{k_i} + o(h), \tag{17}
 \end{aligned}$$

where  $x_{u'} > 0$  and  $k_i \in \{0, 1, 2, \dots\}$  for  $i \in \{1, \dots, \bar{m}\}$ . Here,  $\{\Pi_{u_i u'}(t, h, \mathbf{x})\}_{i \in \{0, 1, \dots, \bar{m}\}}$  is distributed according to the Dirichlet distribution  $\text{Dir}(\{\alpha_{u_i u'}(t, h, \mathbf{x})\}_{i \in \{0, 1, \dots, \bar{m}\}})$  having

$$\alpha_{u_i u'}(t, h, \mathbf{x}) = c \pi_{u_i u'}(t, h, \mathbf{x}) \quad \text{for } i \in \{0, 1, \dots, \bar{m}\}$$

where  $c$  is a finite positive parameter and

$$\begin{aligned}
 &\pi_{u_i u'}(t, h, \mathbf{x}) \\
 &= \begin{cases} \left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} \right) \\ \frac{r_{u_i u'}(t, \mathbf{x})}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{x})} & i \in \{1, \dots, \bar{m}\}, \\ 1 - \sum_{j=1}^{\bar{m}} \pi_{u_j u'}(t, h, \mathbf{x}) & i = 0. \end{cases}
 \end{aligned}$$

The following results hold:

(1) When  $c > 2e^{\sum_{i=1}^{\bar{m}} \int_t^{t+h} r_{u_i u'}(s, \mathbf{x}) ds}$ , for any  $i \in \{1, \dots, \bar{m}\}$ , the infinitesimal mean  $\mu_{u_i u'}^{d\mathbf{X}}(t, \mathbf{x})$  is given by

$$\mu_{u_i u'}^{d\mathbf{X}}(t, \mathbf{x}) = x_{u'} r_{u_i u'}(t, \mathbf{x}) \frac{c}{c - 1},$$

and the infinitesimal variance  $[\sigma_{u_i u'}^{d\mathbf{X}}(t, \mathbf{x})]^2$  is given by

$$[\sigma_{u_i u'}^{d\mathbf{X}}(t, \mathbf{x})]^2 = x_{u'}^2 r_{u_i u'}(t, \mathbf{x}) \frac{c}{(c - 1)(c - 2)}$$



$$+ x_{u'} r_{u_i u'}(t, \mathbf{x}) \frac{c}{c-2}.$$

Then  $\mathbf{X}(t)$  has IOD at  $\mathbf{X}(t) = \mathbf{x}$  with respect to each arrow of  $\{(u_i, u')\}_{i \in \{1, \dots, \bar{m}\}}$ , and  $\mathbf{X}(t)$  has SIOD at  $\mathbf{X}(t) = \mathbf{x}$  for connected incoming arrows  $\{(u_i, u')\}_{i \in \{1, \dots, \bar{m}\}}$ . Furthermore, when  $c > 2e^{\sum_{i=1}^{\bar{m}} \int_i^{i+h} r_{u_i u'}(s, \mathbf{x}) ds}$ , for  $i, j \in \{1, \dots, \bar{m}\}$  and  $i \neq j$ , the infinitesimal covariance  $\sigma_{u_i u', u_j u'}^{d\mathbf{X}}(t, \mathbf{x}) = 0$ .

(2) Denote  $\bar{S}$  as the set of transitions over arrows  $\{(u_i, u')\}_{i \in \{1, \dots, \bar{m}\}}$ , i.e.,

$$\bar{S} := \left\{ k_i : k_i \geq 1 \text{ for } i \in \{1, \dots, \bar{m}\} \right\}. \tag{18}$$

Then the conditional probability that transitions happen over two or more arrows

$$\mathbb{P}\left(\{\Delta_{u_i u'}^{\mathbf{X}}(t, h) = k_i\}_{i \in \{1, \dots, \bar{m}\}}, |\bar{S}| \geq 2 \mid \mathbf{X}(t) = \mathbf{x}\right) = o(h),$$

and the conditional probability that only one transition happens over a single arrow

$$\mathbb{P}\left(\{\Delta_{u_i u'}^{\mathbf{X}}(t, h) = k_i\}_{i \in \{1, \dots, \bar{m}\}}, |\bar{S}| = 1 \mid \mathbf{X}(t) = \mathbf{x}\right) = \sum_{i=1}^{\bar{m}} q_{u_i u'}(t, \mathbf{x}, k_i) h + o(h),$$

where  $|\bar{S}|$  is the cardinality of  $\bar{S}$  and for  $i \in \{1, \dots, \bar{m}\}$

$$q_{u_i u'}(t, \mathbf{x}, k_i) = c \frac{\Gamma(x_{u'} + \sum_{i=1}^{\bar{m}} k_i)}{\Gamma(x_{u'}) \prod_{i=1}^{\bar{m}} \Gamma(k_i + 1)} \frac{\Gamma(x_{u'} + c) \Gamma(k_i)}{\Gamma(x_{u'} + \sum_{i=1}^{\bar{m}} k_i + c)} \times r_{u_i u'}(t, \mathbf{x}). \tag{19}$$

The proof of Theorem 3.2 is postponed to Section A1.2. in the Appendix. We note that in practice, users of Theorem 3.2 can simply treat  $c$  as an unknown parameter through using modern likelihood-based parameter inference algorithms such as Ionides et al. (2015), as illustrated in Sect. 4.

### 4 Application

In this section, we demonstrate our theories, methodologies, and algorithms. A SEIR-type Markov chain model is covered in 1.2. In Sect. 4.1, we provide further details of it. In Sect. 4.2, we use it to conduct measles real data analysis on the same dataset as Bretó et al. (2009) and He et al. (2010), which is a well-tested and publicly accessible dataset, using our methodology and algorithm, for performance comparison.

In Sect. 4.3, we perform a comparative analysis on artificial datasets, generated by perturbing the measles dataset.

#### 4.1 Further details of the motivational application

The standard interpretation of Fig. 1 as a Markov chain having transition rates, conditional on  $\mathbf{x}$ , is given by

$$\begin{aligned} Q(t, (x_B, x_S, \mathbf{x}_{V \setminus \{B, S\}}), (x_B - 1, x_S + 1, \mathbf{x}_{V \setminus \{B, S\}})) &= r_{BS}(t) x_B \mathbb{1}_{\{x_B > 0\}}, \\ Q(t, (x_S, x_E, \mathbf{x}_{V \setminus \{S, E\}}), (x_S - 1, x_E + 1, \mathbf{x}_{V \setminus \{S, E\}})) &= r_{SE}(t, x_I) x_S \mathbb{1}_{\{x_S > 0\}}, \\ Q(t, (x_E, x_I, \mathbf{x}_{V \setminus \{E, I\}}), (x_E - 1, x_I + 1, \mathbf{x}_{V \setminus \{E, I\}})) &= r_{EI} x_E \mathbb{1}_{\{x_E > 0\}}, \\ Q(t, (x_I, x_R, \mathbf{x}_{V \setminus \{I, R\}}), (x_I - 1, x_R + 1, \mathbf{x}_{V \setminus \{I, R\}})) &= r_{IR} x_I \mathbb{1}_{\{x_I > 0\}}, \\ Q(t, (x_S, x_D, \mathbf{x}_{V \setminus \{S, D\}}), (x_S - 1, x_D + 1, \mathbf{x}_{V \setminus \{S, D\}})) &= r_{SD} x_S \mathbb{1}_{\{x_S > 0\}}, \\ Q(t, (x_E, x_D, \mathbf{x}_{V \setminus \{E, D\}}), (x_E - 1, x_D + 1, \mathbf{x}_{V \setminus \{E, D\}})) &= r_{ED} x_E \mathbb{1}_{\{x_E > 0\}}, \\ Q(t, (x_R, x_D, \mathbf{x}_{V \setminus \{R, D\}}), (x_R - 1, x_D + 1, \mathbf{x}_{V \setminus \{R, D\}})) &= r_{RD} x_R \mathbb{1}_{\{x_R > 0\}}, \end{aligned}$$

The time-inhomogeneous transition rate over  $(B, S)$ , denoted as  $r_{BS}(t)$ , is the per-capita rate of recruitment of susceptibles depending on known birth rates obtained via interpolation from birth records. A cohort-entry effect is also considered in calculating  $r_{BS}(t)$ , to reflect the fact that a large cohort of first-year students enters the schools each fall: a fraction  $\theta_c$  of recruits into the susceptible class enter on the school admission day and the remaining fraction  $(1 - \theta_c)$  enter the susceptible class continuously. The TRF of infection  $r_{SE}(t, x_I)$  is defined in (1). Since transmission rates are closely linked to contact rates among children, which are higher during school terms,  $\beta(t)$  reflects the pattern of school terms and holidays, as follows:

$$\beta(t) = \begin{cases} (1 + 2\{1 - p\}\theta_a) \bar{\beta} & \text{during school term,} \\ (1 - 2p\theta_a) \bar{\beta} & \text{during vacation,} \end{cases} \tag{20}$$

where  $p$  is the proportion of the year taken up by school term,  $\bar{\beta}$  is the mean transmission rate, and  $\theta_a$  measures the relative effect of school holidays on transmission. For ease of interpretation,  $\bar{\beta}$  is reparameterized in terms of  $R_0$  which is the annual average basic reproductive ratio, such that  $R_0 = \bar{\beta}/r_{IR}$ , where  $r_{IR}$  is the recovery rate. Here,  $r_{EI}$  is the rate at which exposed individuals become infectious and  $r_{SD} = r_{ED} = r_{ID} = r_{RD}$  denotes a constant per capita death rate.

**Algorithm 2** Euler scheme on generating dynamics having IOD over arrows  $(S, E)$  and  $(S, D)$ , using Algorithm 1.

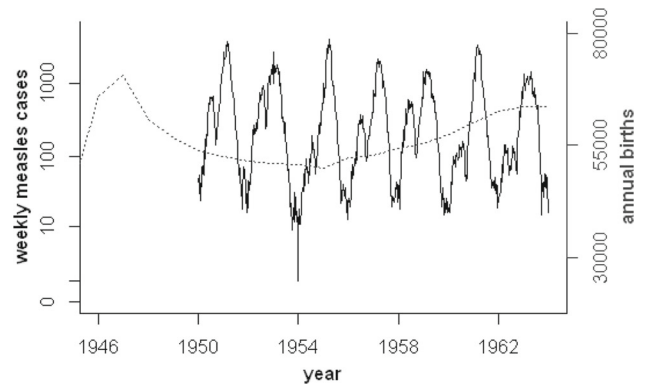
Set the initial value  $\mathbf{X}(0)$  and time interval  $[0, T]$ .  
 Set fixed time increment  $\delta = T/N$  for integer  $N$ ; define  $t_n = n\delta$ .  
 FOR  $n = 0$  to  $N - 1$   
     Generate  $\{\Pi_{S0}, \Pi_{SE}, \Pi_{SD}\}$  according to the Dirichlet distribution  $\text{Dir}(\{\alpha_{S0}, \alpha_{SE}, \alpha_{SD}\})$  having  
      $\alpha_{SE} = c\pi_{SE}$ ,  $\alpha_{SD} = c\pi_{SD}$ , and  $\alpha_{S0} = c - \alpha_{SE} - \alpha_{SD}$ ,  
     where  
      $\pi_{SE} = \left(1 - e^{-\int_n^{n+1} r_{SE}(s, \mathbf{x}) ds - r_{SD}\delta}\right) \frac{r_{SE}(t_n, \mathbf{x})}{r_{SE}(t_n, \mathbf{x}) + r_{SD}}$ ,  
      $\pi_{SD} = \left(1 - e^{-\int_n^{n+1} r_{SE}(s, \mathbf{x}) ds - r_{SD}\delta}\right) \frac{r_{SD}}{r_{SE}(t_n, \mathbf{x}) + r_{SD}}$ .  
     Generate process increments  
      $\{\Delta_{S0}^{\mathbf{X}}, \Delta_{SE}^{\mathbf{X}}, \Delta_{SD}^{\mathbf{X}}\} \sim \text{Multinomial}(X_S(t_n), \{\Pi_{S0}, \Pi_{SE}, \Pi_{SD}\})$   
     where  $\Delta_{S0}^{\mathbf{X}}$  stands for retain individuals.  
     Set  $X_S(t_{n+1}) = \Delta_{S0}^{\mathbf{X}} + \Delta_{BS}^{\mathbf{X}}$   
 END FOR

Our transition rates are taken the same as He et al. (2010) for equidispersed arrows, therefore we use the same Euler approximation. The dynamics over  $(B, S)$  is modelled as an inhomogeneous Poisson process on each step of the Euler scheme. The dynamics over outgoing connected arrows  $\{(S, E), (S, D)\}$  (resp.  $\{(E, I), (E, D)\}, \{(I, R), (I, D)\}$ ) are modeled through multinomial distributions on each step of the Euler scheme, and the dynamic over  $(R, D)$  can be implied through fixed population. The only difference between our approach and that of Bretó et al. (2009), is the modeling of dynamics over  $\{(S, E), (S, D)\}$ . Algorithm 2 is obtained by applying our general Algorithm 1 to this application. In Algorithm 2 the event probabilities in the multinomial distribution are Dirichlet random variables, whereas the approach of Bretó et al. (2009) is adding gamma noise to those probabilities.

**4.2 Comparison using real data**

He et al. (2010) used the over-dispersion methodology (Box 1 on page 280 therein) proposed in Bretó et al. (2009) on analyzing measles epidemics occurring in London during the pre-vaccination era, which is a well-tested and publicly accessible dataset with reported cases from 1950 to 1964. Figure 2 shows the case reports and annual birth rates for London.

In order to conduct fair comparisons, we use the same model setting and data as He et al. (2010). Thus, we fix  $p = 0.7589$  in (20), set the delay from birth to susceptible as 4, and set the mortality rate  $r_{SD} = 1/50 = 0.02$  per year. The unknown model parameters in the SEIR-type Markov chain model covered in Sect. 4, are  $R_0, r_{EI}, r_{IR}, \alpha, \iota, \theta_c$ , and  $\theta_a$ . To calculate the likelihood of the data, a measure-



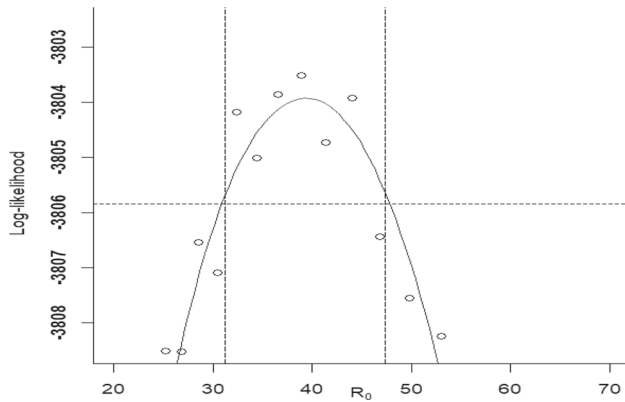
**Fig. 2** Weekly reported measles cases (solid line) and annual births (dotted line) for London

ment model is added to describe the relationship between the latent disease dynamics and the observed case reports. We use the same measurement model as He et al. (2010), which has two more unknown parameters: reporting rate  $\rho$  and dispersion parameter  $\psi$  (see page 281 of He et al. (2010) for a detailed description of this report measurement process). The unknown initializations are  $X_S(0), X_E(0), X_I(0)$ , and  $X_R(0)$ . The unknown IOD model parameters are  $\sigma_{SE}$  of the gamma noise-based approach used in He et al. (2010) and  $c$  of our approach in Algorithm 1. We implemented the same parameter inference algorithm ( Ionides et al. (2015)) as He et al. (2010), via the pomp package ( King et al. (2016)). From Table 1, we can see that with the same number of unknown parameters which indicates the same complexity of inference, our method has better data fitting in terms of a higher ML (Table 3).

There are long-existing concerns about interpreting results generated by the approach proposed in Bretó et al. (2009). The quantity  $R_0$  is central in the epidemiological theory because it has interpretations in terms of many quantities of interest, which include mean age of the first infection, mean susceptible fraction, exponential-phase epidemic growth rate, and vaccination coverage required for eradication. He et al. (2010) obtained  $\text{MLE } R_0 = 56.8$  and the likelihoods over  $R_0$  yielded a 95% confidence interval of (37, 60). Furthermore, Bjørnstad et al. (2002) found an estimate of  $R_0 = 29.9$  for London. Hence, He et al. (2010) gave detailed possible explanations on pages 276 – 278 therein, regarding concerns about the surprisingly high MLE value of  $R_0$ . We obtained  $\text{MLE } R_0 = 34.09$  and a 95% confidence interval of (31.21, 47.37) for London (Fig. 3). Thus, our method not only shows improved statistical fit but also provides a resolution of a previous discrepancy between continuous-time models fitted to time series data and other lines of evidence concerning  $R_0$  for measles.

**Table 1** Comparisons of ML and MLEs using a real dataset

Name	He et al. (2010)	Our method	Name	He et al. (2010)	Our method
ML	<b>- 3804.9</b>	<b>- 3803.2</b>	$\theta_c$	0.56	1
$R_0$	<b>56.8</b>	<b>34.09</b>	$\theta_a$	0.55	0.48
$r_{EI}$	28.9	52.71	$X_S(0)$	0.0297	0.032
$r_{IR}$	30.4	22.88	$X_E(0)$	5.17e-05	6.99e-05
$\alpha$	0.976	1.017	$X_I(0)$	5.14e-05	4.52e-05
$\iota$	2.9	55.08	$X_R(0)$	0.97	0.968
$\rho$	0.488	0.492	$c$	N/A	652.8
$\psi$	0.116	0.118	$\sigma_{SE}$	0.088	N/A



**Fig. 3** Log-likelihood analysis of the basic reproductive ratio,  $R_0$ . The dashed lines construct a 95% confidence interval of (31.21, 47.37) for London

### 4.3 Comparison using artificial datasets

In this subsection, we conduct comparisons using artificial datasets. The real dataset, discussed in the previous subsection, is used to generate two artificial datasets. Artificial dataset 1 is created by perturbing the real dataset from 1950 to 1954, adding one more case to each observation. Artificial dataset 2 is generated by perturbing the same dataset, adding two more cases. We implemented the same parameter inference algorithm ( Ionides et al. (2015)) with the same algorithm setup. Five initial values for the search were generated using five values of  $R_0$  as {10, 15.65, 24.49, 38.34, 60} and the others were set to the best results obtained from the two methods listed in Table 1. Table 2 lists the best two ML and MLE results for each method out of the five initial values. Our method clearly outperforms that of He et al. (2010) in terms of log-likelihood. Furthermore, our method consistently exhibits smaller standard deviations (SD). The parameter inference result for  $R_0$  is much smaller than their estimates when their SD is less than one, which aligns with scientific findings and indicates that the results from He et al. (2010) are less reliable than ours.

**Table 2** Comparisons of ML and MLEs using two artificial datasets. The best two results from each method are listed

He et al. (2010)			Our method		
ML	SD	$R_0$	ML	SD	$R_0$
<i>Artificial Dataset 1</i>					
-1663.26	0.84	41.87	-1653.93	0.22	29.88
-1669.74	0.42	19.78	-1657.44	0.37	25.85
<i>Artificial Dataset 2</i>					
-1659.26	1.66	52.08	-1654.58	0.55	33.47
-1668.88	0.67	45.87	-1661.74	0.31	42.81

## 5 Discussion

Markov processes play a crucial role in stochastic modeling, offering a natural extension of differential equation models used in deterministic modeling. By introducing noise to rates, we achieve a parsimonious approach that allows for flexible mean-variance relationships. While alternative methods like multistate models with expanded state spaces exist, our approach offers distinct advantages: it incorporates a single noise parameter, maintains the Markov property, and preserves the continuous-time structure of the underlying dynamics. These benefits, combined with a good statistical fit to the data, make our model desirable. We demonstrate the effectiveness of our approach in achieving these objectives.

Although simultaneous events are mathematically necessary to achieve flexible mean-variance relationships under the given assumptions, it is important to acknowledge that distinguishing between truly simultaneous and almost-simultaneous events is empirically impossible. In the real world, the process of transitioning between biological or social states, such as becoming infected, lacks a precisely defined exact point in time. Therefore, our focus lies on evaluating whether the modeling structure provides a scientifically meaningful and statistically accurate explanation of observable data. It is these questions that we prioritize in our research.

The topic of  $R_0$  (basic reproduction number) for measles is both intriguing and significant. In a study by Guerra

et al. (2017), more than 16,000 scientific paper abstracts were examined, leading to the identification of 18 studies that reported approximately 60 estimates of  $R_0$  for measles. Among these 18 representative studies, only three were published after the year 2000, focusing on measles data from the years 1950 to 1964 (Table 1, page e422 in Guerra et al. (2017)). Interestingly, all three studies reported relatively high values of  $R_0$ , with He et al. (2010) being one of them. To ensure a fair comparison, our research applies the same data, model settings, and inference algorithm as those employed in He et al. (2010).

The overdispersion approach presented in Bretó et al. (2009) has gained significant popularity; however, it still remains somewhat of a black-box method. This approach lacks a solid mathematical foundation, leaving users uncertain about its conditions of applicability, the extent to which it works, the underlying mechanisms, and its stability. While it may be of interest to examine the limitations of the approach proposed in Bretó et al. (2009), such analysis is beyond the scope of this paper. Instead and importantly, we propose a new theory and methodology that addresses these limitations. In contrast to the previous approach, our method is both explainable and supported by a robust mathematical framework. Firstly, we establish weak yet precise conditions under which our algorithm can be applied more generally. Secondly, through rigorous mathematical derivations, we provide clear insights into how the overdispersion parameter is incorporated into our methodology. Furthermore, our approach can offer improved data fitting and the generation of more interpretable scientific results.

In practice, for complex dynamical systems, not only the joint probability distribution is unknown but also the marginal distributions. A model that is defined using a simulator instead of an analytically tractable characterization of an underlying process, is said to be implicitly defined (Diggle and Gratton 1984). Models for complex dynamic systems that lack analytically tractable transition densities, are sometimes defined implicitly by a computer simulation algorithm. In mathematics and computational science, the Euler method is the commonly used for that purpose. It is a first-order numerical procedure for solving differential equations with a given initial value. Its local error (error per step) is proportional to the square of the step size, and its global error (error at a given time) is proportional to the step size. Our methodology is applicable to general dynamical systems that can be simulated using the Euler scheme, but it is not limited to exclusively using that method. The computational efficiency of the chosen simulation method is the standard for the approach.

Inference approaches that are applicable to implicitly defined models are said to possess the *plug-and-play* property, or alternatively called *equation-free* or *likelihood-free* (Kevrekidis et al. (2004); Xiu et al. (2005); Marjoram et al.

(2003); Sisson et al. (2007)). Modern likelihood-based inference algorithms possessing the *plug-and-play* property (e.g. Ionides et al. (2006, 2015); King et al. (2008)) do not need to evaluate the step-by-step transition densities, and merely require a simulator as input for the dynamic process. Lately, scalable inference algorithms that possess the *plug-and-play* property have been developed for high-dimensional GDMs (e.g. Ning and Ionides (2023); Ionides et al. (2023)). Since our definition of IOD is localized to individual arrows, the dimension of the graph becomes irrelevant. Therefore, our methodology is compatible with those inference algorithms of moderate graph dimension or large graph dimension.

### A1. Proof of Theorems

In this section, we give the proof of Theorem 3.1 in Section A1.1, and the proof of Theorem 3.2 in A1.2. We heavily use properties of the multinomial distribution, the negative multinomial distribution, and the Dirichlet distribution (Table 3). We conduct asymptotic analysis in a sufficiently small time interval  $[t, t + h]$ . We recall that increments over the arrow  $(v, v')$  is defined as

$$\Delta_{vv'}^{\mathbf{X}}(t, h) = N_{vv'}^{\mathbf{X}}(t + h) - N_{vv'}^{\mathbf{X}}(t).$$

We recall that for random variables  $Y_1, Y_2$ , and  $Y_3$ , the law of total variance states that

$$\text{Var}(Y_1) = \mathbb{E}[\text{Var}(Y_1 | Y_2)] + \text{Var}(\mathbb{E}[Y_1 | Y_2]),$$

and the law of total covariance states that

$$\begin{aligned} \text{cov}(Y_1, Y_2) &= \mathbb{E}(\text{cov}(Y_1, Y_2 | Y_3)) \\ &\quad + \text{cov}(\mathbb{E}(Y_1 | Y_3), \mathbb{E}(Y_2 | Y_3)). \end{aligned}$$

#### A1.1 Proof of Theorem 3.1

(1). By (10), for  $i \in \{1, \dots, m\}$

$$\begin{aligned} \mathbb{E}(\Delta_{vv_i}^{\mathbf{X}}(t, h) | \mathbf{X}(t) = \mathbf{x}) &= x_v \mathbb{E}(\Pi_{vv_i}(t, h, \mathbf{x})) + o(h) \\ &= \frac{x_v}{c} \alpha_{vv_i}(t, h, \mathbf{x}) + o(h), \end{aligned}$$

and by the law of total variance

$$\begin{aligned} \text{Var}(\Delta_{vv_i}^{\mathbf{X}}(t, h) | \mathbf{X}(t) = \mathbf{x}) &= x_v^2 \text{Var}(\Pi_{vv_i}(t, h, \mathbf{x})) \\ &\quad + x_v \mathbb{E}(\Pi_{vv_i}(t, h, \mathbf{x})(1 - \Pi_{vv_i}(t, h, \mathbf{x}))) + o(h) \end{aligned}$$

**Table 3** Notation

$G = (V, A)$	$G$ : graph, $V$ : set of vertices, and $A$ : set of arrows, Sect. 2.1.
$S_o$	Set of all source vertices, Sect. 2.1.
$S_i$	Set of all sink vertices, Sect. 2.1.
$X_v(t)$	Stochastic process on vertex $v \in V$ at time $t \in [0, \infty)$ , Sect. 2.1.
$N_{vv'}^{\mathbf{X}}(t)$	Nondecreasing integer-valued jump process over $(v, v')$ , Sect. 2.1.
$\Delta_{vv'}^{\mathbf{X}}(t, h)$	Increments of $N_{vv'}^{\mathbf{X}}(t)$ in time interval $[t, t + h]$ , Sect. 2.2.
$[\sigma_{vv'}^{d\mathbf{X}}(t, \mathbf{x})]^2$	Infinitesimal variance, Def. 2.1.
$\mu_{vv'}^{d\mathbf{X}}(t, \mathbf{x})$	Infinitesimal mean, Def. 2.1.
$D_{vv'}^{d\mathbf{X}}(t, \mathbf{x})$	Infinitesimal dispersion index, Def. 2.1.
$\Upsilon_{vv'}(t, \mathbf{z})$	Transition probability over $(v, v')$ , Sect. 3.
$r_{\cdot\cdot}(t, \mathbf{z})$	Per-capita rate function, Sect. 3.
$\tilde{\pi}_{\cdot\cdot}(t, h, \mathbf{z})$	Transition probability, Sect. 3.
$\pi_{\cdot\cdot}(t, h, \mathbf{z})$	Transition probability after rewriting, Sect. 3.
$\Pi_{\cdot\cdot}(t, h, \mathbf{x})$	Stochastic transition probability, Sect. 3.
$c > 0$	Inverse noise parameter, Sect. 3.
$\alpha_{\cdot\cdot}(t, h, \mathbf{x})$	Parameter in the distribution of $\Pi_{\cdot\cdot}(t, h, \mathbf{x})$ , Sect. 3.
$\beta_{\cdot\cdot}(t, h, \mathbf{x})$	Parameter in the distribution of $\Pi_{\cdot\cdot}(t, h, \mathbf{x})$ , Sect. 3.
$Q(t, \mathbf{x}, \mathbf{x}')$	Transition rate function, Eqn. (3).
$q_{vv'}(t, \mathbf{x}, k)$	Transition rate function, Eqn. (9).
$S$	Set defined in Eqn. (11).
$\bar{S}$	Set defined in Eqn. (18).

$$= (x_v^2 - x_v) \frac{\alpha_{vv'_i}(t, h, \mathbf{x})(c - \alpha_{vv'_i}(t, h, \mathbf{x}))}{c^2(c + 1)} + \frac{x_v}{c} \alpha_{vv'_i}(t, h, \mathbf{x}) - \frac{x_v}{c^2} (\alpha_{vv'_i}(t, h, \mathbf{x}))^2 + o(h).$$

**Plugging**

$$\alpha_{vv'_i}(t, h, \mathbf{x}) = c \left( 1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{x}) ds} \right) \frac{r_{vv'_i}(t, \mathbf{x})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{x})}$$

into the above two equations, we have

$$\mathbb{E}(\Delta_{vv'_i}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}) = x_v \left( 1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{x}) ds} \right) \times \frac{r_{vv'_i}(t, \mathbf{x})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{x})} + o(h)$$

and

$$\text{Var}(\Delta_{vv'_i}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}) = (x_v^2 - x_v)(c + 1)^{-1}$$

$$\begin{aligned} & \times \left( 1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{x}) ds} \right) \frac{r_{vv'_i}(t, \mathbf{x})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{x})} \\ & \times \left( 1 - \left( 1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{x}) ds} \right) \frac{r_{vv'_i}(t, \mathbf{x})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{x})} \right) \\ & + x_v \left( 1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{x}) ds} \right) \frac{r_{vv'_i}(t, \mathbf{x})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{x})} \\ & - x_v \left( \left( 1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{x}) ds} \right) \frac{r_{vv'_i}(t, \mathbf{x})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{x})} \right)^2 \\ & + o(h) \\ & := \mathcal{A}_1 + \mathcal{A}_2 - \mathcal{A}_3 + o(h). \end{aligned} \tag{A.1}$$

We need to ensure that the above variance is positive. Since  $x_v \geq 1$ ,  $c > 0$ , and  $r_{vv'_i}(t, \mathbf{x}) > 0$ , we have

$$0 < \left( 1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{x}) ds} \right) < 1$$

and  $0 < \frac{r_{vv'_i}(t, \mathbf{x})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{x})} < 1$ ,

and then the first term  $\mathcal{A}_1$  is always positive. Since

$$\left( 1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{x}) ds} \right) \frac{r_{vv'_i}(t, \mathbf{x})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{x})}$$

$$> \left( \left( 1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{x}) ds} \right) \frac{r_{vv'_i}(t, \mathbf{x})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{x})} \right)^2,$$

we have  $\mathcal{A}_2 - \mathcal{A}_3 > 0$ . Thus, when  $h$  is sufficiently small, we have

$$\text{Var}(\Delta_{vv'_i}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}) > 0.$$

Note that when  $h$  is sufficiently small, we also have  $\mathbb{E}(\Delta_{vv'_i}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}) > 0$ . Applying L'Hôpital's rule,

$$\begin{aligned} \mu_{vv'_i}^{d\mathbf{X}}(t, \mathbf{x}) &= x_v r_{vv'_i}(t, \mathbf{x}), \\ [\sigma_{vv'_i}^{d\mathbf{X}}(t, \mathbf{x})]^2 &= (1 + (x_v - 1)(c + 1)^{-1}) x_v r_{vv'_i}(t, \mathbf{x}). \end{aligned}$$

Then we can see that for any  $i \in \{1, \dots, m\}$ , when  $x_v > 1$ , the infinitesimal dispersion index  $D_{vv'_i}^{d\mathbf{X}}(t, \mathbf{x}) > 1$ , i.e.,  $\mathbf{X}(t)$  has IOD at  $\mathbf{X}(t) = \mathbf{x}$  with respect to arrow  $(v, v'_i) \in A$ ; when  $x_v = 1$ , the infinitesimal dispersion index  $D_{vv'_i}^{d\mathbf{X}}(t, \mathbf{x}) = 1$ , i.e.,  $\mathbf{X}(t)$  has IED at  $\mathbf{X}(t) = \mathbf{x}$  with respect to arrow  $(v, v'_i) \in A$ . By Definition 2.2, when  $x_v > 1$ ,  $\mathbf{X}(t)$  has SIOD at  $\mathbf{X}(t) = \mathbf{x}$  over connected outgoing arrows  $\{(v, v'_i)\}_{i \in \{1, \dots, m\}}$ ; when  $x_v = 1$ ,  $\mathbf{X}(t)$  has SIED at  $\mathbf{X}(t) = \mathbf{x}$  over connected outgoing arrows  $\{(v, v'_i)\}_{i \in \{1, \dots, m\}}$ .

By the law of total covariance, for  $i, j \in \{1, \dots, m\}$  and  $i \neq j$ , we have

$$\begin{aligned} &\text{Cov}[\Delta_{vv'_i}^{\mathbf{X}}(t, h), \Delta_{vv'_j}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}] \\ &= \text{Cov}[\mathbb{E}[\Delta_{vv'_i}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}, \{\Pi_{vv'_i}(t, h, \mathbf{x})\}_{i \in \{0, \dots, m\}}], \\ &\mathbb{E}[\Delta_{vv'_j}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}, \{\Pi_{vv'_i}(t, h, \mathbf{x})\}_{i \in \{0, \dots, m\}}]] \\ &+ \mathbb{E}[\text{Cov}[\Delta_{vv'_i}^{\mathbf{X}}(t, h), \Delta_{vv'_j}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}, \\ &\quad \{\Pi_{vv'_i}(t, h, \mathbf{x})\}_{i \in \{0, \dots, m\}}]] \\ &= x_v^2 \text{Cov}[\Pi_{vv'_i}(t, h, \mathbf{x}), \Pi_{vv'_j}(t, h, \mathbf{x})] \\ &\quad - \mathbb{E}[x_v \Pi_{vv'_i}(t, h, \mathbf{x}) \Pi_{vv'_j}(t, h, \mathbf{x})] + o(h) \\ &= (x_v^2 - x_v) \text{Cov}[\Pi_{vv'_i}(t, h, \mathbf{x}), \Pi_{vv'_j}(t, h, \mathbf{x})] \\ &\quad - x_v \mathbb{E}[\Pi_{vv'_i}(t, h, \mathbf{x})] \mathbb{E}[\Pi_{vv'_j}(t, h, \mathbf{x})] + o(h) \\ &= -(x_v^2 - x_v) \frac{\alpha_{vv'_i}(t, h, \mathbf{x}) \alpha_{vv'_j}(t, h, \mathbf{x})}{c^2(c + 1)} \\ &\quad - x_v \frac{\alpha_{vv'_i}(t, h, \mathbf{x}) \alpha_{vv'_j}(t, h, \mathbf{x})}{c^2} + o(h) \\ &= -\frac{(x_v^2 - x_v)}{c + 1} \left( 1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{x}) ds} \right)^2 \\ &\quad \times \frac{r_{vv'_i}(t, \mathbf{x}) r_{vv'_j}(t, \mathbf{x})}{(\sum_{j=1}^m r_{vv'_j}(t, \mathbf{x}))^2} \end{aligned}$$

$$\begin{aligned} &- x_v \left( 1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{x}) ds} \right)^2 \\ &\quad \times \frac{r_{vv'_i}(t, \mathbf{x}) r_{vv'_j}(t, \mathbf{x})}{(\sum_{j=1}^m r_{vv'_j}(t, \mathbf{x}))^2} + o(h). \end{aligned} \tag{A.2}$$

By Definition 2.3, for  $i, j \in \{1, \dots, m\}$  and  $i \neq j$ ,

$$\sigma_{vv'_i, vv'_j}^{d\mathbf{X}}(t, \mathbf{x}) = 0.$$

(2). By (10), with  $\mathbf{y} = (y_0, y_1, \dots, y_m)$ , we have

$$\begin{aligned} &\mathbb{P}(\{\Delta_{vv'_i}^{\mathbf{X}}(t, h) = k_i\}_{i \in \{0, \dots, m\}} \mid \mathbf{X}(t) = \mathbf{x}) \\ &= \int_0^1 \frac{\Gamma(x_v + 1)}{\prod_{i=0}^m \Gamma(k_i + 1)} \prod_{i=0}^m (y_i)^{k_i} \\ &\quad \times \left\{ \frac{\Gamma(\sum_{i=0}^m \alpha_{vv'_i}(t, h, \mathbf{x}))}{\prod_{i=0}^m \Gamma(\alpha_{vv'_i}(t, h, \mathbf{x}))} \prod_{i=0}^m y_i^{\alpha_{vv'_i}(t, h, \mathbf{x}) - 1} \right\} dy \\ &\quad + o(h) \\ &= \frac{\Gamma(x_v + 1)}{\prod_{i=0}^m \Gamma(k_i + 1)} \frac{\Gamma(\sum_{i=0}^m \alpha_{vv'_i}(t, h, \mathbf{x}))}{\prod_{i=0}^m \Gamma(\alpha_{vv'_i}(t, h, \mathbf{x}))} \\ &\quad \times \int \prod_{i=0}^m (y_i)^{k_i + \alpha_{vv'_i}(t, h, \mathbf{x}) - 1} dy + o(h) \\ &= \frac{\Gamma(x_v + 1)}{\prod_{i=0}^m \Gamma(k_i + 1)} \frac{\Gamma(\sum_{i=0}^m \alpha_{vv'_i}(t, h, \mathbf{x}))}{\prod_{i=0}^m \Gamma(\alpha_{vv'_i}(t, h, \mathbf{x}))} \\ &\quad \times \frac{\prod_{i=0}^m \Gamma(k_i + \alpha_{vv'_i}(t, h, \mathbf{x}))}{\Gamma(\sum_{i=0}^m k_i + \sum_{i=0}^m \alpha_{vv'_i}(t, h, \mathbf{x}))} + o(h) \\ &= \frac{\Gamma(x_v + 1)}{\prod_{i=0}^m \Gamma(k_i + 1)} \frac{\Gamma(c)}{\prod_{i=0}^m \Gamma(\alpha_{vv'_i}(t, h, \mathbf{x}))} \\ &\quad \times \frac{\prod_{i=0}^m \Gamma(k_i + \alpha_{vv'_i}(t, h, \mathbf{x}))}{\Gamma(x_v + c)} + o(h). \end{aligned} \tag{A.3}$$

Recalling that for  $i \in \{1, \dots, m\}$

$$\begin{aligned} \alpha_{vv'_i}(t, h, \mathbf{x}) &= c \left( 1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{x}) ds} \right) \\ &\quad \times \frac{r_{vv'_i}(t, \mathbf{x})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{x})}, \end{aligned}$$

by Taylor series we have

$$\alpha_{vv'_i}(t, h, \mathbf{x}) = c r_{vv'_i}(t, \mathbf{x}) h + o(h). \tag{A.4}$$

Hence, for  $k_i \geq 1$  and  $i \in \{1, \dots, m\}$ ,

$$\Gamma(k_i + \alpha_{vv'_i}(t, h, \mathbf{x}))$$

$$\begin{aligned}
 &= (k_i + \alpha_{vv'_i}(t, h, \mathbf{x}) - 1) \cdots (2 + \alpha_{vv'_i}(t, h, \mathbf{x}) - 1) \\
 &\quad \times \alpha_{vv'_i}(t, h, \mathbf{x}) \cdot \Gamma(\alpha_{vv'_i}(t, h, \mathbf{x})) \\
 &= (k_i - 1) \cdots (2 - 1) \cdot \alpha_{vv'_i}(t, h, \mathbf{x}) \cdot \Gamma(\alpha_{vv'_i}(t, h, \mathbf{x})) \\
 &\quad + o(\alpha_{vv'_i}(t, h, \mathbf{x})) \\
 &= \Gamma(k_i)\alpha_{vv'_i}(t, h, \mathbf{x})\Gamma(\alpha_{vv'_i}(t, h, \mathbf{x})) + o(\alpha_{vv'_i}(t, h, \mathbf{x})).
 \end{aligned} \tag{A.5}$$

Furthermore, we have that

$$\begin{aligned}
 &\Gamma(k_0 + \alpha_{vv'_0}(t, h, \mathbf{x})) \\
 &= \Gamma\left(x_v - \sum_{i=1}^m k_{i \in |A|} + c - \sum_{i=1}^m \alpha_{vv'_i}(t, h, \mathbf{x})\right) \\
 &= \left(x_v - \sum_{i=1}^m k_{i \in |A|} - 1 + c - \sum_{i=1}^m \alpha_{vv'_i}(t, h, \mathbf{x})\right) \\
 &\quad \times \cdots \times \left(c - \sum_{i=1}^m \alpha_{vv'_i}(t, h, \mathbf{x})\right) \\
 &\quad \times \Gamma(\alpha_{vv'_0}(t, h, \mathbf{x})) \\
 &= \left[\left(x_v - \sum_{i=1}^m k_{i \in |A|} - 1 + c\right) \cdots c + o\left(\sum_{i=1}^m \alpha_{vv'_i}(t, h, \mathbf{x})\right)\right] \\
 &\quad \times \Gamma(\alpha_{vv'_0}(t, h, \mathbf{x})) \\
 &= \left[\frac{\Gamma(x_v - \sum_{i=1}^m k_{i \in |A|} + c)}{\Gamma(c)} + o\left(\sum_{i=1}^m \alpha_{vv'_i}(t, h, \mathbf{x})\right)\right] \Gamma(\alpha_{vv'_0}(t, h, \mathbf{x})).
 \end{aligned} \tag{A.6}$$

Plugging (A.4), (A.5), and (A.6) into (A.3), we can see that

$$\mathbb{P}\left(\{\Delta_{vv'_i}^{\mathbf{X}}(t, h) = k_i\}_{i \in \{0, 1, \dots, m\}}, |\mathcal{S}| \geq 2 \mid \mathbf{X}(t) = \mathbf{x}\right) = o(h)$$

and

$$\begin{aligned}
 &\mathbb{P}\left(\{\Delta_{vv'_i}^{\mathbf{X}}(t, h) = k_i\}_{i \in \{0, 1, \dots, m\}}, |\mathcal{S}| = 1 \mid \mathbf{X}(t) = \mathbf{x}\right) \\
 &= \sum_{i=1}^m q_{vv'_i}(t, \mathbf{x}, k_i)h + o(h),
 \end{aligned}$$

where  $\mathcal{S}$  is the set defined in (11),  $|\mathcal{S}|$  is the cardinality of  $\mathcal{S}$ , and for  $i \in \{1, \dots, m\}$

$$q_{vv'_i}(t, \mathbf{x}, k_i) = c \binom{x_v}{k_i} \frac{\Gamma(k_i)\Gamma(x_v - k_i + c)}{\Gamma(x_v + c)} r_{vv'_i}(t, \mathbf{x}).$$

The proof is complete.

**Remark A1.1** It is worth noting that if we use the approximation

$$\left(1 - e^{-\sum_{j=1}^m \int_t^{t+h} r_{vv'_j}(s, \mathbf{x}) ds}\right) \frac{r_{vv'_i}(t, \mathbf{x})}{\sum_{j=1}^m r_{vv'_j}(t, \mathbf{x})} \approx r_{vv'_i}(t, \mathbf{x})h,$$

equation (A.1) would be

$$\begin{aligned}
 &\text{Var}(\Delta_{vv'_i}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}) \\
 &= (x_v^2 - x_v)(c + 1)^{-1} r_{vv'_j}(t, \mathbf{x})h \left(1 - r_{vv'_j}(t, \mathbf{x})h\right) \\
 &\quad + x_v r_{vv'_j}(t, \mathbf{x})h - x_v r_{vv'_j}^2(t, \mathbf{x})h^2 + o(h) \\
 &= x_v r_{vv'_j}(t, \mathbf{x})h \left(1 - r_{vv'_j}(t, \mathbf{x})h\right) \\
 &\quad \times \left[(x_v - 1)(c + 1)^{-1} + 1\right] + o(h).
 \end{aligned}$$

For  $r_{vv'_j}(t, \mathbf{x})$  defined in equation (5), we did not assume  $r_{vv'_j}(t, \mathbf{x})h \in (0, 1)$  uniformly for all  $t$  and  $\mathbf{x}$ . Hence, this approximated form cannot guarantee the positivity of  $\text{Var}(\Delta_{vv'_i}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x})$ .

### A1.2. Proof of Theorem 3.2

(1). By (17), for  $i \in \{1, \dots, \bar{m}\}$ , with  $\mathbf{y} = (y_0, y_1, \dots, y_{\bar{m}})$ , we have

$$\begin{aligned}
 &\mathbb{E}(\Delta_{u_i u'}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}) \\
 &= x_{u'} \mathbb{E}\left(\left(\prod_{u_0 u'}(t, h, \mathbf{x})\right)^{-1} \prod_{u_i u'}(t, h, \mathbf{x})\right) + o(h) \\
 &= x_{u'} \int_0^1 y_0^{-1} y_i \frac{\Gamma(\sum_{i=0}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}))}{\prod_{i=0}^{\bar{m}} \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}))} \\
 &\quad \times \prod_{i=0}^{\bar{m}} y_i^{\alpha_{u_i u'}(t, h, \mathbf{x})-1} d\mathbf{y} + o(h) \\
 &= \int_0^1 y_0^{\alpha_{u_0 u'}(t, h, \mathbf{x})-2} y_i^{\alpha_{u_i u'}(t, h, \mathbf{x})} \prod_{j=1}^{i-1} y_j^{\alpha_{u_j u'}(t, h, \mathbf{x})-1} \\
 &\quad \times \prod_{j=i+1}^{\bar{m}} y_j^{\alpha_{u_j u'}(t, h, \mathbf{x})-1} d\mathbf{y} \\
 &\quad \times x_{u'} \frac{\Gamma(\sum_{i=0}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}))}{\prod_{i=0}^{\bar{m}} \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}))} + o(h) \\
 &= x_{u'} \frac{\Gamma(\sum_{i=0}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}))}{\prod_{i=0}^{\bar{m}} \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}))} \\
 &\quad \times \frac{\prod_{j=1}^{i-1} \Gamma(\alpha_{u_j u'}(t, h, \mathbf{x})) \prod_{j=i+1}^{\bar{m}} \Gamma(\alpha_{u_j u'}(t, h, \mathbf{x}))}{\Gamma(\sum_{i=0}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}))} \\
 &\quad \times x_{u'} \Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x}) - 1) \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}) + 1) + o(h) \\
 &= x_{u'} \frac{\Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x}) - 1) \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}) + 1)}{\Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x})) \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}))} + o(h) \\
 &= x_{u'} \frac{\alpha_{u_i u'}(t, h, \mathbf{x})}{\alpha_{u_0 u'}(t, h, \mathbf{x}) - 1} + o(h).
 \end{aligned} \tag{A.7}$$

By (17) and the law of total variance, we have

$$\begin{aligned} & \text{Var}(\Delta_{u_i u'}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}) \\ &= x_{u'}^2 \text{Var}((\Pi_{u_0 u'}(t, h, \mathbf{x}))^{-1} \Pi_{u_i u'}(t, h, \mathbf{x})) \\ & \quad + x_{u'} \mathbb{E}[(\Pi_{u_0 u'}(t, h, \mathbf{x}))^{-2} (\Pi_{u_i u'}(t, h, \mathbf{x}))^2] \\ & \quad + x_{u'} \mathbb{E}[(\Pi_{u_0 u'}(t, h, \mathbf{x}))^{-1} \Pi_{u_i u'}(t, h, \mathbf{x})] + o(h) \\ &= (x_{u'}^2 + x_{u'}) \mathbb{E}[(\Pi_{u_0 u'}(t, h, \mathbf{x}))^{-2} (\Pi_{u_i u'}(t, h, \mathbf{x}))^2] \\ & \quad - x_{u'}^2 \mathbb{E}[(\Pi_{u_0 u'}(t, h, \mathbf{x})) (\Pi_{u_i u'}(t, h, \mathbf{x}))^2] \\ & \quad + x_{u'} \mathbb{E}[(\Pi_{u_0 u'}(t, h, \mathbf{x}))^{-1} \Pi_{u_i u'}(t, h, \mathbf{x})] + o(h) \\ &= (x_{u'}^2 + x_{u'}) \frac{\Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x}) - 2) \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}) + 2)}{\Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x})) \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}))} \\ & \quad - x_{u'}^2 \left( \frac{\Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x}) - 1) \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}) + 1)}{\Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x})) \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}))} \right)^2 \\ & \quad + x_{u'} \left( \frac{\Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x}) - 1) \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}) + 1)}{\Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x})) \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}))} \right) \\ & \quad + o(h) \\ &= (x_{u'}^2 + x_{u'}) \frac{(\alpha_{u_i u'}(t, h, \mathbf{x}) + 1) \alpha_{u_i u'}(t, h, \mathbf{x})}{(\alpha_{u_0 u'}(t, h, \mathbf{x}) - 1) (\alpha_{u_0 u'}(t, h, \mathbf{x}) - 2)} \\ & \quad - x_{u'}^2 \left( \frac{\alpha_{u_i u'}(t, h, \mathbf{x})}{\alpha_{u_0 u'}(t, h, \mathbf{x}) - 1} \right)^2 + x_{u'} \left( \frac{\alpha_{u_i u'}(t, h, \mathbf{x})}{\alpha_{u_0 u'}(t, h, \mathbf{x}) - 1} \right) \\ & \quad + o(h). \end{aligned} \tag{A.8}$$

Plugging

$$\begin{aligned} \alpha_{u_i u'}(t, h, \mathbf{x}) &= c \left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} \right) \\ & \quad \times \frac{r_{u_i u'}(t, \mathbf{x})}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{x})} \end{aligned} \tag{A.9}$$

and

$$\begin{aligned} \alpha_{u_0 u'}(t, h, \mathbf{x}) &= c - \sum_{i=1}^m \alpha_{u_i u'}(t, h, \mathbf{x}) \\ &= c e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} \end{aligned} \tag{A.10}$$

into equations (A.7) and (A.8), we have

$$\begin{aligned} & \mathbb{E}(\Delta_{u_i u'}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}) \\ &= x_{u'} \frac{c \left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} \right) \frac{r_{u_i u'}(t, \mathbf{x})}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{x})}}{c e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} - 1} \\ & \quad + o(h) \end{aligned} \tag{A.11}$$

and

$$\text{Var}(\Delta_{u_i u'}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x})$$

$$\begin{aligned} &= (x_{u'}^2 + x_{u'}) \\ & \quad \times \frac{c \left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} \right) \frac{r_{u_i u'}(t, \mathbf{x})}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{x})} + 1}{\left( c e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} - 1 \right) \left( c e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} - 2 \right)} \\ & \quad \times c \left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} \right) \frac{r_{u_i u'}(t, \mathbf{x})}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{x})} \\ & \quad - x_{u'}^2 \left( \frac{c \left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} \right) \frac{r_{u_i u'}(t, \mathbf{x})}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{x})}}{c e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} - 1} \right)^2 \\ & \quad + x_{u'} \left( \frac{c \left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} \right) \frac{r_{u_i u'}(t, \mathbf{x})}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{x})}}{c e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} - 1} \right) \\ & \quad + o(h). \end{aligned} \tag{A.12}$$

Note that when  $h$  is sufficiently small,  $c > 2e^{\sum_{i=1}^{\bar{m}} \int_t^{t+h} r_{u_i u'}(s, \mathbf{x}) ds}$  suffices that  $\mathbb{E}(\Delta_{u_i u'}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}) > 0$ , and together with the fact that

$$\begin{aligned} & \frac{\left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} \right) \frac{r_{u_i u'}(t, \mathbf{x})}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{x})} + 1}{c e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} - 2} \\ & > \frac{c \left( 1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} \right) \frac{r_{u_i u'}(t, \mathbf{x})}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{x})}}{c e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} - 1} \end{aligned}$$

we also have  $\text{Var}(\Delta_{u_i u'}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}) > 0$ .

By L'Hôpital's rule, equations (A.11) and (A.12), and Definition 2.1, we have

$$\begin{aligned} \mu_{u_i u'}^{d\mathbf{X}}(t, \mathbf{x}) &= x_{u'} r_{u_i u'}(t, \mathbf{x}) \frac{c}{c - 1}, \\ [\sigma_{u_i u'}^{d\mathbf{X}}(t, \mathbf{x})]^2 &= x_{u'}^2 r_{u_i u'}(t, \mathbf{x}) \frac{c}{(c - 1)(c - 2)} \\ & \quad + x_{u'} r_{u_i u'}(t, \mathbf{x}) \frac{c}{c - 2}. \end{aligned}$$

When  $c > 2e^{\sum_{i=1}^{\bar{m}} \int_t^{t+h} r_{u_i u'}(s, \mathbf{x}) ds}$ , since

$$x_{u'} r_{u_i u'}(t, \mathbf{x}) \frac{c}{c - 2} > x_{u'} r_{u_i u'}(t, \mathbf{x}) \frac{c}{c - 1} > 0$$

and

$$x_{u'}^2 r_{u_i u'}(t, \mathbf{x}) \frac{c}{(c - 1)(c - 2)} > 0,$$

we have  $D_{u_i u'}^{d\mathbf{X}}(t, \mathbf{x}) > 1$ , i.e.,  $\mathbf{X}(t)$  has IOD at  $\mathbf{X}(t) = \mathbf{x}$  with respect to each arrow in  $\{(u_i, u')\}_{i \in \{1, \dots, \bar{m}\}}$ . By Definition 2.2,



$\mathbf{X}(t)$  has SIOD at  $\mathbf{X}(t) = \mathbf{x}$  over connected incoming arrows  $\{(u_i, u')\}_{i \in \{1, \dots, \bar{m}\}}$ .

By the law of total covariance, for  $i, j \in \{1, \dots, \bar{m}\}$  and  $i \neq j$ , we have

$$\begin{aligned} & \text{Cov}\left[\Delta_{u_i u'}^{\mathbf{X}}(t, h), \Delta_{u_j u'}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}\right] \\ &= \text{Cov}\left[\mathbb{E}\left[\Delta_{u_i u'}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}, \{\Pi_{u_i u'}(t, h, \mathbf{x})\}_{i \in \{0, 1, \dots, \bar{m}\}}\right], \right. \\ & \quad \times \mathbb{E}\left[\Delta_{u_j u'}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}, \{\Pi_{u_j u'}(t, h, \mathbf{x})\}_{j \in \{0, 1, \dots, \bar{m}\}}\right] \\ & \quad \left. + \mathbb{E}\left[\text{Cov}\left[\Delta_{u_i u'}^{\mathbf{X}}(t, h), \Delta_{u_j u'}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}, \right. \right. \right. \\ & \quad \quad \left. \left. \left. \{\Pi_{u_i u'}(t, h, \mathbf{x})\}_{i \in \{0, 1, \dots, \bar{m}\}}\right] + o(h)\right] \right. \\ &= x_u^2 \text{Cov}\left[\left(\Pi_{u_0 u'}(t, h, \mathbf{x})\right)^{-1} \Pi_{u_i u'}(t, h, \mathbf{x}), \right. \\ & \quad \left. \left(\Pi_{u_0 u'}(t, h, \mathbf{x})\right)^{-1} \Pi_{u_j u'}(t, h, \mathbf{x})\right] \\ & \quad + \mathbb{E}\left[x_{u'}\left(\Pi_{u_0 u'}(t, h, \mathbf{x})\right)^{-2} \Pi_{u_i u'}(t, h, \mathbf{x}) \Pi_{u_j u'}(t, h, \mathbf{x})\right] + o(h) \\ &= x_u^2 \mathbb{E}\left[\left(\Pi_{u_0 u'}(t, h, \mathbf{x})\right)^{-2} \Pi_{u_i u'}(t, h, \mathbf{x}) \Pi_{u_j u'}(t, h, \mathbf{x})\right] \\ & \quad - x_u^2 \mathbb{E}\left[\left(\Pi_{u_0 u'}(t, h, \mathbf{x})\right)^{-1} \Pi_{u_i u'}(t, h, \mathbf{x})\right] \\ & \quad \times \mathbb{E}\left[\left(\Pi_{u_0 u'}(t, h, \mathbf{x})\right)^{-1} \Pi_{u_j u'}(t, h, \mathbf{x})\right] \\ & \quad + x_{u'} \mathbb{E}\left[\left(\Pi_{u_0 u'}(t, h, \mathbf{x})\right)^{-2} \Pi_{u_i u'}(t, h, \mathbf{x}) \Pi_{u_j u'}(t, h, \mathbf{x})\right] + o(h) \\ &= (x_u^2 + x_{u'}) \mathbb{E}\left[\left(\Pi_{u_0 u'}(t, h, \mathbf{x})\right)^{-2} \Pi_{u_i u'}(t, h, \mathbf{x}) \Pi_{u_j u'}(t, h, \mathbf{x})\right] \\ & \quad - x_u^2 \mathbb{E}\left[\left(\Pi_{u_0 u'}(t, h, \mathbf{x})\right)^{-1} \Pi_{u_i u'}(t, h, \mathbf{x})\right] \\ & \quad \times \mathbb{E}\left[\left(\Pi_{u_0 u'}(t, h, \mathbf{x})\right)^{-1} \Pi_{u_j u'}(t, h, \mathbf{x})\right] + o(h). \end{aligned} \tag{A.13}$$

Note that

$$\begin{aligned} & \mathbb{E}\left[\left(\Pi_{u_0 u'}(t, h, \mathbf{x})\right)^{-2} \Pi_{u_i u'}(t, h, \mathbf{x}) \Pi_{u_j u'}(t, h, \mathbf{x})\right] \\ &= \int_0^1 y_0^{-2} y_i y_j \frac{\Gamma(\sum_{i=0}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}))}{\prod_{i=0}^{\bar{m}} \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}))} \\ & \quad \times \prod_{i=0}^{\bar{m}} \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x})) \prod_{i=0}^{\bar{m}} y_i^{\alpha_{u_i u'}(t, h, \mathbf{x})-1} dy + o(h) \\ &= \frac{\Gamma(\sum_{i=0}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}))}{\prod_{i=0}^{\bar{m}} \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}))} \\ & \quad \times \int_0^1 y_0^{\alpha_{u_0 u'}(t, h, \mathbf{x})-3} y_i^{\alpha_{u_i u'}(t, h, \mathbf{x})} y_j^{\alpha_{u_j u'}(t, h, \mathbf{x})} \\ & \quad \times \prod_{l \in \{1, \dots, \bar{m}\} \setminus \{i, j\}} y_l^{\alpha_{u_l u'}(t, h, \mathbf{x})-1} dy + o(h) \\ &= \Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x}) - 2) \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}) + 1) \Gamma(\alpha_{u_j u'}(t, h, \mathbf{x}) + 1) \\ & \quad \times \frac{\Gamma(\sum_{i=0}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x})) \prod_{l \in \{1, \dots, \bar{m}\} \setminus \{i, j\}} \Gamma(\alpha_{u_l u'}(t, h, \mathbf{x}))}{\prod_{i=0}^{\bar{m}} \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x})) \Gamma(\sum_{i=0}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}))} + o(h) \\ &= \frac{\Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x}) - 2) \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}) + 1) \Gamma(\alpha_{u_j u'}(t, h, \mathbf{x}) + 1)}{\Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x})) \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x})) \Gamma(\alpha_{u_j u'}(t, h, \mathbf{x}))} \\ & \quad + o(h) \end{aligned}$$

$$\begin{aligned} &= \frac{\alpha_{u_i u'}(t, h, \mathbf{x}) \alpha_{u_j u'}(t, h, \mathbf{x})}{(\alpha_{u_0 u'}(t, h, \mathbf{x}) - 1)(\alpha_{u_0 u'}(t, h, \mathbf{x}) - 2)} \\ & \quad + o(h). \end{aligned} \tag{A.14}$$

Plugging equations (A.9) and (A.10) into (A.14), we have

$$\begin{aligned} & \mathbb{E}\left[\left(\Pi_{u_0 u'}(t, h, \mathbf{x})\right)^{-2} \Pi_{u_i u'}(t, h, \mathbf{x}) \Pi_{u_j u'}(t, h, \mathbf{x})\right] \\ &= \frac{c^2 \left(1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds}\right)^2 \frac{r_{u_i u'}(t, \mathbf{x}) r_{u_j u'}(t, \mathbf{x})}{\left(\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{x})\right)^2}}{\left(ce^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} - 1\right) \left(ce^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} - 2\right)} \\ & \quad + o(h). \end{aligned} \tag{A.15}$$

Plugging (A.15) into (A.13), we have

$$\begin{aligned} & \text{Cov}\left[\Delta_{u_i u'}^{\mathbf{X}}(t, h), \Delta_{u_j u'}^{\mathbf{X}}(t, h) \mid \mathbf{X}(t) = \mathbf{x}\right] \\ &= (x_u^2 + x_{u'}) \\ & \quad \times \frac{c^2 \left(1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds}\right)^2 \frac{r_{u_i u'}(t, \mathbf{x}) r_{u_j u'}(t, \mathbf{x})}{\left(\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{x})\right)^2}}{\left(ce^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} - 1\right) \left(ce^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} - 2\right)} \\ & \quad - x_u^2 \frac{c \left(1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds}\right) \frac{r_{u_i u'}(t, \mathbf{x})}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{x})}}{\left(ce^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} - 1\right)} \\ & \quad - x_{u'} \frac{c \left(1 - e^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds}\right) \frac{r_{u_j u'}(t, \mathbf{x})}{\sum_{j=1}^{\bar{m}} r_{u_j u'}(t, \mathbf{x})}}{\left(ce^{-\sum_{j=1}^{\bar{m}} \int_t^{t+h} r_{u_j u'}(s, \mathbf{x}) ds} - 1\right)} \\ & \quad + o(h). \end{aligned} \tag{A.16}$$

By Definition 2.3, for  $i, j \in \{1, \dots, \bar{m}\}$  and  $i \neq j$ ,

$$\sigma_{u_i u', u_j u'}^{d\mathbf{X}}(t, \mathbf{x}) = 0.$$

(2). By (17), with  $\mathbf{y} = (y_0, y_1, \dots, y_{\bar{m}})$ , we have

$$\begin{aligned} & \mathbb{P}\left(\{\Delta_{u_i u'}^{\mathbf{X}}(t, h) = k_i\}_{i \in \{1, \dots, \bar{m}\}} \mid \mathbf{X}(t) = \mathbf{x}\right) \\ &= \int_0^1 \frac{\Gamma(x_{u'} + \sum_{i=1}^{\bar{m}} k_i)}{\Gamma(x_{u'}) \prod_{i=1}^{\bar{m}} \Gamma(k_i + 1)} [y_0]^{x_{u'}} \\ & \quad \times \prod_{i=1}^{\bar{m}} [y_i]^{k_i} \frac{\Gamma(\sum_{i=0}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}))}{\prod_{i=0}^{\bar{m}} \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}))} \\ & \quad \times \prod_{i=0}^{\bar{m}} y_i^{\alpha_{u_i u'}(t, h, \mathbf{x})-1} dy + o(h) \\ &= \frac{\Gamma(x_{u'} + \sum_{i=1}^{\bar{m}} k_i)}{\Gamma(x_{u'}) \prod_{i=1}^{\bar{m}} \Gamma(k_i + 1)} \frac{\Gamma(\sum_{i=0}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}))}{\prod_{i=0}^{\bar{m}} \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}))} \\ & \quad \times \int_0^1 [y_0]^{x_{u'} + \alpha_{u_0 u'}(t, h, \mathbf{x})-1} \prod_{i=1}^{\bar{m}} [y_i]^{k_i + \alpha_{u_i u'}(t, h, \mathbf{x})-1} dy \end{aligned}$$

$$\begin{aligned}
 &+ o(h) \\
 &= \frac{\Gamma(x_{u'} + \sum_{i=1}^{\bar{m}} k_i)}{\Gamma(x_{u'}) \prod_{i=1}^{\bar{m}} \Gamma(k_i + 1)} \frac{\Gamma(\sum_{i=0}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}))}{\prod_{i=0}^{\bar{m}} \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}))} \\
 &\quad \times \frac{\Gamma(x_{u'} + \alpha_{u_0 u'}(t, h, \mathbf{x})) \prod_{i=1}^{\bar{m}} \Gamma(k_i + \alpha_{u_i u'}(t, h, \mathbf{x}))}{\Gamma(x_{u'} + \sum_{i=1}^{\bar{m}} k_i + \sum_{i=0}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}))} \\
 &+ o(h) \\
 &= \frac{\Gamma(x_{u'} + \sum_{i=1}^{\bar{m}} k_i)}{\Gamma(x_{u'}) \prod_{i=1}^{\bar{m}} \Gamma(k_i + 1)} \frac{\Gamma(c)}{\prod_{i=0}^{\bar{m}} \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x}))} \\
 &\quad \times \frac{\Gamma(x_{u'} + \alpha_{u_0 u'}(t, h, \mathbf{x})) \prod_{i=1}^{\bar{m}} \Gamma(k_i + \alpha_{u_i u'}(t, h, \mathbf{x}))}{\Gamma(x_{u'} + \sum_{i=1}^{\bar{m}} k_i + c)} \\
 &+ o(h). \tag{A.17}
 \end{aligned}$$

Recalling that for  $i \in \{1, \dots, \bar{m}\}$

$$\begin{aligned}
 &\alpha_{u_i u'}(t, h, \mathbf{x}) \\
 &= c \left( 1 - e^{-\sum_{i=1}^{\bar{m}} \int_t^{t+h} r_{u_i u'}(s, \mathbf{x}) ds} \right) \frac{r_{u_i u'}(t, \mathbf{x})}{\sum_{i=1}^{\bar{m}} r_{u_i u'}(t, \mathbf{x})},
 \end{aligned}$$

by Taylor series we have

$$\alpha_{u_i u'}(t, h, \mathbf{x}) = cr_{u_i u'}(t, \mathbf{x})h + o(h). \tag{A.18}$$

Hence, for  $k_i \geq 1$  and  $i \in \{1, \dots, \bar{m}\}$ ,

$$\begin{aligned}
 &\Gamma(k_i + \alpha_{u_i u'}(t, h, \mathbf{x})) \\
 &= (k_i + \alpha_{u_i u'}(t, h, \mathbf{x}) - 1) \\
 &\quad \times \dots \times (2 + \alpha_{u_i u'}(t, h, \mathbf{x}) - 1) \alpha_{u_i u'}(t, h, \mathbf{x}) \\
 &\quad \times \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x})) \\
 &= (k_i - 1) \dots (2 - 1) \cdot \alpha_{u_i u'}(t, h, \mathbf{x}) \\
 &\quad \times \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x})) + o(\alpha_{u_i u'}(t, h, \mathbf{x})) \\
 &= \Gamma(k_i) \alpha_{u_i u'}(t, h, \mathbf{x}) \Gamma(\alpha_{u_i u'}(t, h, \mathbf{x})) + o(\alpha_{u_i u'}(t, h, \mathbf{x})). \tag{A.19}
 \end{aligned}$$

Furthermore, we have

$$\begin{aligned}
 &\Gamma(x_{u'} + \alpha_{u_0 u'}(t, h, \mathbf{x})) \\
 &= \left( x_{u'} - 1 + c - \sum_{i=1}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}) \right) \\
 &\quad \times \dots \times \left( c - \sum_{i=1}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}) \right) \\
 &\quad \times \Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x})) \\
 &= \left[ (x_{u'} - 1 + c) \dots c + \mathcal{O} \left( \sum_{i=1}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}) \right) \right] \\
 &\quad \times \Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x}))
 \end{aligned}$$

$$\begin{aligned}
 &= \left[ \frac{\Gamma(x_{u'} + c)}{\Gamma(c)} + \mathcal{O} \left( \sum_{i=1}^{\bar{m}} \alpha_{u_i u'}(t, h, \mathbf{x}) \right) \right] \\
 &\quad \times \Gamma(\alpha_{u_0 u'}(t, h, \mathbf{x})). \tag{A.20}
 \end{aligned}$$

Plugging (A.18), (A.19) and (A.20) into (A.17), we can see that

$$\mathbb{P} \left( \{ \Delta_{u_i u'}^{\mathbf{X}}(t, h) = k_i \}_{i \in \{1, \dots, \bar{m}\}}, |\bar{S}| \geq 2 \mid \mathbf{X}(t) = \mathbf{x} \right) = o(h)$$

and

$$\begin{aligned}
 &\mathbb{P} \left( \{ \Delta_{u_i u'}^{\mathbf{X}}(t, h) = k_i \}_{i \in \{1, \dots, \bar{m}\}}, |\bar{S}| = 1 \mid \mathbf{X}(t) = \mathbf{x} \right) \\
 &= \sum_{i=1}^{\bar{m}} q_{u_i u'}(t, \mathbf{x}, k_i) h + o(h),
 \end{aligned}$$

where  $\bar{S}$  is the set defined in (18),  $|\bar{S}|$  is the cardinality of  $\bar{S}$ , and

$$\begin{aligned}
 &q_{u_i u'}(t, \mathbf{x}, k_i) \\
 &= c \frac{\Gamma(x_{u'} + \sum_{i=1}^{\bar{m}} k_i)}{\Gamma(x_{u'}) \prod_{i=1}^{\bar{m}} \Gamma(k_i + 1)} \frac{\Gamma(x_{u'} + c) \Gamma(k_i)}{\Gamma(x_{u'} + \sum_{i=1}^{\bar{m}} k_i + c)} \\
 &\quad \times r_{u_i u'}(t, \mathbf{x}).
 \end{aligned}$$

**Acknowledgements** The authors would like to thank two anonymous reviewers and the Editors for their very constructive comments and efforts on this lengthy work, which greatly improved the quality of this paper. This research project was partially supported by NSF grant DMS-1761603, NIH grant 1R21AI180492-01 and the Individual Research Grant at Texas A&M University.

**Author Contributions** Ning wrote the main manuscript text. All authors reviewed and revised the manuscript.

**Declarations**

**Conflict of interest** Ning Ning serves as an Associate Editor for Statistics and Computing.

**References**

Albrecher, H., Beirlant, J., Teugels, J.F.: Reinsurance: Actuarial and Statistical Aspects. Wiley (2017)

Berk, R., MacDonald, J.M.: Overdispersion and Poisson regression. J. Quant. Criminol. **24**(3), 269–284 (2008)

Bjørnstad, O.N., Finkenstädt, B.F., Grenfell, B.T.: Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. Ecol. Monogr. **72**(2), 169–184 (2002)

Bretó, C.: On the infinitesimal dispersion of multivariate Markov counting systems. Stat. Probab. Lett. **82**(4), 720–725 (2012)

Bretó, C.: Co-jumps and Markov counting systems in random environments. In: Contemporary Approaches and Methods in Fundamental Mathematics and Mechanics, pp. 277–292. Springer (2021)

- Bretó, C., Ionides, E.L.: Compound Markov counting processes and their applications to modeling infinitesimally over-dispersed systems. *Stoch. Process. Appl.* **121**(11), 2571–2591 (2011)
- Bretó, C., He, D., Ionides, E.L., King, A.A.: Time series analysis via mechanistic models. *Ann. Appl. Stat.* **3**(1), 319–348 (2009)
- Brown, T.C., Hamza, K., Xia, A.: On the variance to mean ratio for random variables from Markov chains and point processes. *J. Appl. Probab.* **35**, 303–312 (1998)
- Chen, S., Shojaie, A., Witten, D.M.: Network reconstruction from high-dimensional ordinary differential equations. *J. Am. Stat. Assoc.* **112**(520), 1697–1707 (2017)
- Diggle, P.J., Gratton, R.J.: Monte Carlo methods of inference for implicit statistical models. *J. R. Stat. Soc.: Ser. B (Methodol.)* **46**(2), 193–212 (1984)
- Ding, K., Ning, N.: Markov chain approximation and measure change for time-inhomogeneous stochastic processes. *Appl. Math. Comput.* **392**, 125732 (2021)
- Gao, D., Sendova, K.P.: Applications of the classical compound Poisson model with claim sizes following a compound distribution. *Probab. Eng. Inform. Sci.* **37**, 1–30 (2022)
- Gillespie, J.H.: The molecular clock may be an episodic clock. *Proc. Natl. Acad. Sci.* **81**(24), 8009–8013 (1984)
- Guerra, F.M., Bolotin, S., Lim, G., Heffernan, J., Deeks, S.L., Li, Y., Crowcroft, N.S.: The basic reproduction number ( $R_0$ ) of measles: a systematic review. *Lancet. Infect. Dis* **17**(12), e420–e428 (2017)
- He, D., Ionides, E.L., King, A.A.: Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *J. R. Soc. Interface* **7**(43), 271–283 (2010)
- Ionides, E.L., Bretó, C., King, A.A.: Inference for nonlinear dynamical systems. *Proc. Natl. Acad. Sci.* **103**(49), 18438–18443 (2006)
- Ionides, E.L., Bhadra, A., Atchadé, Y., King, A.: Iterated filtering. *Ann. Stat.* **39**(3), 1776–1802 (2011)
- Ionides, E.L., Nguyen, D., Atchadé, Y., Stoev, S., King, A.A.: Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proc. Natl. Acad. Sci.* **112**(3), 719–724 (2015)
- Ionides, E.L., Asfaw, K., Park, J., King, A.A.: Bagged filters for partially observed spatiotemporal systems. *J. Am. Stat. Assoc.* (2021). <https://doi.org/10.1080/01621459.2021.1974867>
- Ionides, E.L., Ning, N., Wheeler, J.: An iterated block particle filter for inference on coupled dynamic systems with shared and unit-specific parameters. *arXiv preprint arXiv:2206.03837*. *Statistica Sinica* (forthcoming), (2023)
- Jayaraman, R., Matis, T.I.: Batch arrivals and service-single station queues. *Wiley Encyclopedia of Operations Research and Management Science* (2010)
- Katzfuss, M., Stroud, J.R., Wikle, C.K.: Ensemble Kalman methods for high-dimensional hierarchical dynamic space-time models. *J. Am. Stat. Assoc.* **115**(530), 866–885 (2020)
- Kevrekidis, I.G., William Gear, C., Hummer, G.: Equation-free: the computer-aided analysis of complex multiscale systems. *AICHE J.* **50**(7), 1346–1355 (2004)
- King, A.A., Ionides, E.L., Pascual, M., Bouma, M.J.: Inapparent infections and cholera dynamics. *Nature* **454**(7206), 877–880 (2008)
- King, A.A., Nguyen, D., Ionides, E.L.: Statistical inference for partially observed Markov processes via the R package pomp. *J. Stat. Softw.* **69**(12) (2016)
- Klugman, S.A., Panjer, H.H., Willmot, G.E.: *Loss Models: Further Topics*. Wiley (2013)
- Krak, T., De Bock, J., Siebes, A.: Efficient computation of updated lower expectations for imprecise continuous-time hidden Markov chains. In: *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, pp. 193–204. *Proceedings of Machine Learning Research* (2017)
- Li, Y., Sendova, K.P.: A surplus process involving a compound Poisson counting process and applications. *Commun. Stat.-Theory Methods* **49**(13), 3238–3256 (2020)
- Maraghi, F.A., Madan, K.C., Darby-Dowman, K.: Batch arrival queueing system with random breakdowns and Bernoulli schedule server vacations having general vacation time distribution. *Int. J. Inf. Manage. Sci.* **20**(1), 55–70 (2009)
- Marjoram, P., Molitor, J., Plagnol, V., Tavaré, S.: Markov chain Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci.* **100**(26), 15324–15328 (2003)
- Ning, N., Ionides, E.L.: Iterated block particle filter for high-dimensional parameter learning: beating the curse of dimensionality. *J. Mach. Learn. Res.* **24**(82), 1–76 (2023)
- Ning, N., Jing, W.: Well-posedness and stability analysis of two classes of generalized stochastic volatility models. *SIAM J. Financ. Math.* **12**(1), 79–109 (2021)
- Parthasarathy, K.R.: *Probability measures on metric spaces*, volume 352. American Mathematical Soc. (2005)
- Ramsay, J., Hooker, G.: *Dynamic Data Analysis*. Springer (2017)
- Sendova, K.P., Minkova, L.D.: Poisson-logarithmic risk process and applications. *Comptes rendus de l'Académie bulgare des Sciences* **71** (8) (2018)
- Sendova, K.P., Minkova, L.D.: Introducing the non-homogeneous compound-birth process. *Stochastics* **92**(5), 814–832 (2020)
- Sisson, S.A., Fan, Y., Tanaka, M.M.: Sequential Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci.* **104**(6), 1760–1765 (2007)
- Xiu, D., Kevrekidis, I.G., Ghanem, R.: An equation-free, multiscale approach to uncertainty quantification. *Comput. Sc. Eng.* **7**(3), 16–23 (2005)
- Zhang, H., Li, B.: Characterizations of discrete compound Poisson distributions. *Commun. Stat.-Theory Methods* **45**(22), 6789–6802 (2016)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.